Blind Adaptive Estimation of KLT Basis Vectors

Carlos E. Davila

Electrical Engineering Department, Southern Methodist University
Dallas, Texas 75275-0338

E-mail: cd@seas.smu.edu
Phone: (214) 768-3197
Fax: (214) 768-3573

Abstract

Several algorithms for estimating the basis vectors used in the Karhunen-Loeve Transform (KLT) are described. The algorithms are “blind” in the sense that they utilize minimal information about the data vector being encoded. This eliminates the need to repeatedly encode and transmit the KLT basis vectors in data compression applications.

1 Introduction

The Karhunen-Loeve Transform (KLT) is known to be the optimum transform for signal compression [1, 2]. Unfortunately, the KLT basis functions, which are the eigenvectors of the data correlation matrix, are data dependent. Hence the basis functions must also be encoded and transmitted which reduces compression and leads to increased data rates. For this reason, the KLT has found limited use in data compression applications. Let

\[ x_n = \begin{bmatrix} x(nN - 1) & x(nN - 2) & \cdots & x(N(n - 1)) \end{bmatrix}^T \quad (1) \]

be the \( N \)-dimensional signal frame to be encoded. We assume that \( x_n \) has correlation matrix \( R = E [x_n x_n^T] \) having rank \( r \leq N \). This means that \( x_n \) can be represented as a linear combination of the eigenvectors of \( R \) given by \( q^1, q^2, \ldots, q^r \), corresponding to eigenvalues \( \lambda^1 \geq \lambda^2 \geq \ldots \geq \lambda^r > 0 \), respectively. Let \( Q = \begin{bmatrix} q^1 & q^2 & \cdots & q^r \end{bmatrix} \) be an \( N \times r \) matrix whose columns are the KLT basis vectors (eigenvectors of \( R \)). The transform coefficients, given by \( y_n = Q^T x_n \), can then be quantized as \( \hat{y}_n \), encoded, and transmitted. If the receiver has knowledge of the basis vectors \( Q \), \( x_n \) can be recovered as \( \hat{x}_n = Q \hat{y}_n \). If the signal \( x(n) \) is statistically stationary then the eigenvectors need only be estimated and transmitted once, which would not lead to much loss of compression, however in practice, the eigenstructure of most signals tends to vary considerably over time. Hence the eigenvectors of \( R \) need to be constantly retransmitted which is why the KLT is not often used. In this paper, we give a method of determining the basis vectors for the KLT directly from the KLT coefficients given only very limited knowledge of \( x(n) \). This eliminates the need to retransmit the KLT basis vectors.
2 Tracking KLT Basis Vectors

Blind estimation of the KLT basis vectors can be accomplished using ideas from the subspace tracking literature. Let $\hat{R}_n$ be an estimate of $E[x_n x_n^T]$ at that is updated using,

$$\hat{R}_n = \gamma \hat{R}_{n-1} + x_n x_n^T$$  (2)

where $0 < \gamma < 1$. Let

$$\hat{Q}_n = \begin{bmatrix} \hat{q}_n^1 & \hat{q}_n^2 & \cdots & \hat{q}_n^r \end{bmatrix}$$  (3)

and $\hat{\Lambda}_n = diag(\hat{\lambda}_n^1, \hat{\lambda}_n^2, \ldots, \hat{\lambda}_n^r)$ be estimates of the eigenvectors and eigenvalues, respectively, of $\hat{R}_n$. Then $\hat{R}_n \approx \gamma \hat{Q}_{n-1} \hat{\Lambda}_{n-1} \hat{Q}_{n-1}^T + x_n x_n^T$. The eigenvector estimates can be updated as follows:

1. Solve the generalized eigenvalue problem

$$FW_n = GW_n \Pi_n$$  (4)

where $F = \hat{Q}_n^T \left( \gamma \hat{Q}_{n-1} \hat{\Lambda}_{n-1} \hat{Q}_{n-1}^T + x_n x_n^T \right) \hat{Q}_n$, $G = \hat{Q}_n^T \hat{Q}_n$, and $W_n$ and the diagonal matrix $\Pi_n$ are the respective generalized eigenvectors and eigenvalues. The matrix $\overline{Q}_n = \begin{bmatrix} \hat{Q}_{n-1} & v_n \end{bmatrix}$ has dimension $N \times (r + 1)$ and $v_n$ is a search direction vector.

2. Update the eigenvector estimates as

$$\hat{Q}_n = \overline{Q}_n W_n(1: r)$$  (5)

where $W_n(1: r)$ are the eigenvectors corresponding to the maximum $r$ eigenvalues in $\Pi_n$.

3. The eigenvalue estimates are updated as $\hat{\Lambda}_n = \Pi_n(1 : r, 1 : r)$. 
\[
\begin{align*}
\hat{Q}_0 &= I_N(:, 1:r) \\
\text{for } n = 1, 2, \ldots \quad &
\begin{cases}
\bar{Q}_n = [Q_{n-1} \quad v_n ] \\
y_n = \hat{Q}_{n-1}^T x_n \\
y_n = [y_n^T \ x_n^T v_n ]^T \\
y_n = \Delta(y_n) \\
\text{transmit } \hat{y}_n \text{ to receiver} \\
F = \gamma \bar{Q}_n^T \hat{Q}_{n-1} \hat{\Lambda}_{n-1} \bar{Q}_{n-1}^T \bar{Q}_n + \hat{y}_n \hat{y}_n^T \\
G = \bar{Q}_n^T \bar{Q}_n \\
solve \ FW_n = GW_n \Pi_n \\
\hat{Q}_n = \bar{Q}_n W_n(1:r)
\end{cases}
end
\end{align*}
\]

Table 1: Basic algorithm AKLT1 for blind estimation of KLT basis vectors. The search direction vector \(v_n\) is assumed known to both the sender and receiver for each \(n\); \(\Delta(\cdot)\) is a quantizer.

If, in the above algorithm, the search direction is set to \(v_n = x_n\), then the algorithm is a standard subspace averaging algorithm used for subspace tracking [3, 4]. Note that if we treat the columns of \(\hat{Q}_{n-1}\) as the KLT basis vectors, then the KLT coefficients are contained in the first \(r\) elements of \(\bar{Q}_n^T x_n\) (see (4)), hence the algorithm never explicitly uses \(x_n\). We will show that if \(v_n\) is a white noise vector, independent of \(x_n\), then the eigenvectors of \(\hat{R}_n\) can still be tracked. This implies that the above algorithm can be run by both the sender and the receiver concurrently using the same initial conditions. If the search direction vectors are known to both the sender and the receiver, then the receiver can also track the KLT basis vectors having only knowledge of the KLT coefficients and the additional scalar coefficient, \(v_n^T x_n\).

In this context, the algorithm is “blind” since the receiver requires no explicit knowledge of the signal \(x(n)\) to track the KLT basis vectors. Table 1 lists the algorithm (AKLT1), \(I_N\) is the \(N \times N\) identity matrix.

### 3 Algorithm Convergence

The key to analyzing the algorithm’s convergence is the fact that the generalized eigenvalue problem in the algorithm is equivalent to an orthogonal projection. This is stated in the
following theorem.

**Theorem 1:** Let $\tilde{R}_n \equiv \gamma_{n-1} \hat{\Lambda}_{n-1} \hat{Q}_{n-1}^T + x_n x_n^T$, let $P_{\hat{Q}_n}$ be a projection matrix onto $\mathcal{R} \left( \hat{Q}_n \right)$ where $\hat{Q}_n$ are the eigenvector estimates in (5), and let $S$ be any $r$-dimensional subspace of the range of $\mathcal{Q}_n$, denoted $\mathcal{R}(\mathcal{Q}_n)$. The projection matrix onto $S$ is denoted by $P_{\mathcal{S}}$. Then the subspace $\mathcal{R} \left( \hat{Q}_n \right)$ maximizes $tr P_{\mathcal{S}} \tilde{R}_n P_{\mathcal{S}}$ over all subspaces $S$, i.e.

$$\max_S \left[ tr P_{\mathcal{S}} \tilde{R}_n P_{\mathcal{S}} \right] = tr P_{\hat{Q}_n} \tilde{R}_n P_{\hat{Q}_n}$$

**Proof:** First, we shall seek to maximize $tr U^T \tilde{R}_n U$ over all $N \times r$ matrices $U$ having orthonormal columns in $\mathcal{R}(\mathcal{Q}_n)$. The maximizing matrix $U^*$ can be found by construction. The algorithm in the previous section finds the unit-norm vector $\hat{q}_1^1$ in $\mathcal{R}(\mathcal{Q}_n)$ which maximizes $\hat{q}_1^1 \tilde{R}_n \hat{q}_1^1$. It simultaneously finds the unit-norm vector $\hat{q}_2^1 \perp \hat{q}_1^1$ in $\mathcal{R}(\mathcal{Q}_n)$ maximizing $\hat{q}_2^1 \tilde{R}_n \hat{q}_2^1$, and proceeds in this way to find orthonormal vectors $\hat{q}_k^1, k = 3, \ldots, r$ in $\mathcal{R}(\mathcal{Q}_n)$ which maximize $\hat{q}_k^1 \tilde{R}_n \hat{q}_k^1$ such that $\hat{q}_k^1$ is orthogonal to the span of $\hat{q}_1^1, \ldots, \hat{q}_{k-1}^1$. This is a consequence of the min-max property of the eigenvalues of the symmetric matrix $\hat{Q}_n^T \hat{R}_n \hat{Q}_n$ [Householder, 1964], where $\hat{Q}_n$ is an $N \times r$ matrix having orthonormal columns and the same range as $\mathcal{Q}_n$. Therefore the matrix $U^*$ which maximizes $tr U^T \tilde{R}_n U$ is the estimated eigenvector matrix $\hat{Q}_n = \left[ \begin{array}{ccc} \hat{q}_1^1 & \hat{q}_2^1 & \cdots & \hat{q}_r^1 \end{array} \right]$. Since $\hat{Q}_n$ contains orthonormal vectors, it follows that $\hat{Q}_n^T \hat{Q}_n = I$. Using a well-known property of the trace operation, the maximized trace can be written as

$$tr \hat{Q}_n^T \tilde{R}_n \hat{Q}_n = tr \hat{R}_n \hat{Q}_n \hat{Q}_n^T$$

$$= tr \hat{R}_n P_{\hat{Q}_n}$$

$$= tr P_{\hat{Q}_n} \hat{R}_n P_{\hat{Q}_n}$$

$$= tr P_{\hat{Q}_n} \hat{R}_n P_{\hat{Q}_n}$$
where the last equality has used the idempotent property of projection matrices. This establishes the Theorem. Q.E.D.

Given that the quantity in (6) is the maximum over all projection matrices $P_S$, it follows that $\text{tr} P^\perp_{\tilde{Q}_n} \tilde{R}_n P^\perp_{\tilde{Q}_n}$ is the minimum. Hence, a suitable measure of the algorithm’s convergence is

$$\epsilon(n) \equiv \text{tr} E \left[ P^\perp_{\tilde{Q}_n} \left( \gamma \hat{Q}_{n-1} \hat{A}_{n-1} \hat{Q}^T_{n-1} + x_n x_n^T \right) P^\perp_{\tilde{Q}_n} \right]$$

(7)

where $P^\perp_{\tilde{Q}_n} = I - P_{\tilde{Q}_n}$. We can write

$$P^\perp_{\tilde{Q}_n} = \begin{cases} I - P_{\tilde{Q}_{n-1}} & \frac{P^\perp_{\tilde{Q}_{n-1}} v_n v_n^T P^\perp_{\tilde{Q}_{n-1}}}{v_n^T P^\perp_{\tilde{Q}_{n-1}} v_n} \\ P^\perp_{\tilde{Q}_{n-1}} & \frac{P^\perp_{\tilde{Q}_{n-1}} v_n v_n^T P^\perp_{\tilde{Q}_{n-1}} v_n}{v_n^T P^\perp_{\tilde{Q}_{n-1}} v_n} \end{cases}$$

(8)

Hence, using the fact that $v_n$ and $x_n$ are independent,

$$\epsilon(n) = \text{tr} E \left[ P^\perp_{\tilde{Q}_{n-1}} x_n x_n^T - \frac{P^\perp_{\tilde{Q}_{n-1}} v_n v_n^T P^\perp_{\tilde{Q}_{n-1}}}{v_n^T P^\perp_{\tilde{Q}_{n-1}} v_n} x_n x_n^T \right]$$

(10)
where $A_n^k = \frac{P_{\hat{Q}_n-1}^\perp q^k q^T P_{\hat{Q}_n-1}^\perp}{q^T q_{\hat{Q}_n-1}^k q^k}$, $B_n = P_{\hat{Q}_n-1}^\perp$, and the expectation is assumed to be conditioned on $x_m, v_m, m = 0, \ldots, n - 1$. We can also write

$$
\epsilon(n) = trE \left[ P_{\hat{Q}_n}^\perp x_n x_n^T \right]
= E \left[ trP_{\hat{Q}_n}^\perp \right] R
= \sum_{k=1}^{r} \lambda^k q^k E \left[ P_{\hat{Q}_n}^\perp q^k \right]
= \sum_{k=1}^{r} \lambda^k q^k E \left[ P_{\hat{Q}_n}^\perp q^k \right] \equiv \sum_{k=1}^{r} \epsilon^k(n)
$$

(11)

From the last equality in (11), we see that $\epsilon(n) \to 0$ as $\hat{Q}_n \to Q$. Using (10) and (11), each mode of the error measure can be written as

$$
\epsilon^k(n) = \epsilon^k(n-1) \left( 1 - E \left[ \frac{v_n^T A_n^k v_n}{v_n^T B_n v_n} \right] \right), \quad k = 1, \ldots, r
$$

(12)

Hence, the reduction in each mode of $\epsilon(n)$ at each time step is determined by $E \left[ v_n^T A_n^k v_n / v_n^T B_n v_n \right]$, which has an upper bound of 1. This expectation can be evaluated using the method developed by Bershad for analyzing the normalized LMS algorithm [5]. To simplify the notation, we henceforth replace $A_n^k$ and $B_n$ with $A$ and $B$, respectively.

**Theorem 2:** If we assume that $v_n$ is a zero mean, unit variance Gaussian white noise vector then

$$
E \left[ \frac{v_n^T A v_n}{v_n^T B v_n} \right] = \int \frac{dg}{d\beta} d\beta \bigg|_{\beta=0}
$$

(13)

where

$$
g(\beta) \equiv \frac{1}{(2\pi)^{N/2}} \int \frac{v_n^T A v_n}{v_n^T B v_n} \exp \left\{ -\beta v_n^T B v_n \right\} \exp \left\{ -\frac{v_n^T v_n}{2} \right\} dv
$$

(14)

$$
\frac{dg}{d\beta} = -|C|^{1/2} a^T C a, \quad C = (I + 2\beta B)^{-1}, \quad a = P_{\hat{Q}_n-1}^\perp q_i / \sqrt{q_i^T P_{\hat{Q}_n-1}^\perp q_i}, \quad i = 1, \ldots, r.
$$
Proof: Since \( v_n \) is assumed to be Gaussian white noise:

\[
E \left[ v_n^T A v_n \right] = \frac{1}{(2\pi)^{N/2}} \int \frac{v_n^T A v_n}{v_n^T B v_n} \exp \left\{ -\frac{v_n^T v_n}{2} \right\} \, dv \tag{15}
\]

The derivative of \( g(\beta) \) leads to an integral which is more readily evaluated

\[
\frac{dg}{d\beta} = -\frac{1}{(2\pi)^{N/2}} \int v_n^T A v_n \exp \left\{ -\beta v_n^T B v_n \right\} \exp \left\{ -\frac{v_n^T v_n}{2} \right\} \, dv \tag{16}
\]

where \( C^{-1} = (I + 2\beta B) \). The quantity in the square brackets is the expectation of \( v_n^T A v_n \)
where \( v_n \) is a zero-mean Gaussian vector with covariance matrix \( C \). Letting

\[
a = P_{Q_{n-1}}^\perp q^i / \sqrt{q^iT P_{Q_{n-1}}^\perp q^i}, i = 1, \ldots, r \tag{17}
\]

it follows that \( A = aa^T \) is a constant rank-1 matrix so that \( E \left[ v_n^T A v_n \right] = trAE \left[ v_n v_n^T \right] = trAC = a^T Ca \) hence

\[
\frac{dg}{d\beta} = -|C|^{1/2} a^T Ca \tag{18}
\]

Therefore

\[
E \left[ \frac{v_n^T A v_n}{v_n^T B v_n} \right] = \int \frac{dg}{d\beta} d\beta \bigg|_{\beta=0} + D \tag{19}
\]

where the integration constant \( D = 0 \) [5].

**Theorem 3:** \( \rho(N, r) \equiv E \left[ \frac{v_n^T A v_n}{v_n^T B v_n} \right] = 1/(N - r) \).

**Proof:** We note that:

\[
C^{-1} = I_N + 2\beta B = I_N + 2\beta \left( I_N - P_{Q_{n-1}} \right) \tag{20}
\]

\[
= (1 + 2\beta) I_N - 2\beta P_{Q_{n-1}}
\]
Let $\xi_i, i = 1, \ldots, N$ be the eigenvalues of $C^{-1}$, then since the eigenvalues of projection matrices are one and zero,

$$
\xi^i = \begin{cases} 
1, & i = 1, \ldots, r \\
1 + 2\beta, & i = r + 1, \ldots, N 
\end{cases}
$$

(21)

Therefore

$$
C = \hat{Q}_{n-1}\hat{Q}_{n-1}^T + (1 + 2\beta)^{-1} \left( I_N - \hat{Q}_{n-1}\hat{Q}_{n-1}^T \right)
$$

(22)

and from the definition of $a$, we have

$$
a^T Ca = (1 + 2\beta)^{-1}
$$

(23)

From (21) it follows that $|C^{-1}|^{1/2} = (1 + 2\beta)^{(N-r)/2}$, and hence,

$$
|C|^{1/2} = (1 + 2\beta)^{-(N-r)/2}
$$

(24)

therefore, from (18) (23), and (24), we get

$$
\frac{dg}{d\beta} = -(1 + 2\beta)^{(r-N)/2-1}
$$

(25)

which when integrated and evaluated at $\beta = 0$, gives $\rho(N, r) = 1/(N - r)$. Q.E.D.

We note that when $r = N - 1$, $\rho(N, r) = 1$ since the columns of $\overline{q}_n$ (the search space) span the entire space $\mathbb{R}^N$. Also note that in order to maintain the same convergence speed, the quantity $N - r$ should be held fixed. This means that attempting to use a large value of $N$ and a small value of $r$, which increases compression, leads to a reduction in conversion speed. To look at the effects on algorithm convergence when the vector $\overline{y}_n$ is quantized, let $\hat{y}_n = \overline{y}_n + e_n$ where $e_n$ is a quantization error vector. If the quantizer is fine enough, then $e_n$ can be modeled as a white noise vector [2]. Moreover, the quantizer can be chosen
so that each element of $e_n$ has equal power. Under these conditions, quantizing the KLT coefficient vector $\tilde{y}_n$ is equivalent to adding a white noise vector $z_n$, to the data vector, $x_n$. This can easily be seen by noting that $\hat{y}_n = Q_n^T(x_n + z_n) = \tilde{y}_n + e_n$, where $E[z_n z_n^T] = \sigma_z^2 I_N$ and assuming that $v_n^TQ_{n-1} \approx 0$ and $\|v_n\| = 1$. Since adding a diagonal matrix to a square symmetric matrix ($R$ in this case) has no effect on its eigenvectors, quantizing *per se* does not affect the KLT basis vectors. However adding white noise to the data vector does affect the tracking dynamics and error in the subspace tracking algorithm. This will be analyzed more fully in a future paper.

4 Increasing Convergence Speed

The reduction in the eigenvector estimation error $\epsilon(n)$ can be slow for a white noise search direction, particularly for larger values of $N - r$. One way of improving convergence speed is with a codebook search. We define a codebook as an array $V$ of dimension $N \times M$ which contains $M$ candidate search directions. Both the sender and the receiver have the same codebook, the sender searches the codebook for the search direction which maximizes $tr\Pi_n(1 : r, 1 : r)$ and uses that search direction to update the KLT basis vectors. The sender then sends the KLT coefficients and the codebook index for the optimum search direction to the receiver who then updates his own KLT basis vectors as seen in Table 1. This approach has been found to greatly increase the convergence and tracking rate of the algorithm; the tradeoff is having to transmit the codebook index. A detailed listing of the codebook search algorithm AKLT2 is given in Table 2.

5 Complete Search Space Algorithm

Algorithms AKLT1 and AKLT2 assume $r$, the dimension of the signal subspace, is known. Some signals of practical interest have a signal subspace dimension which changes with time. Moreover, during sudden changes in the statistics of the signal, the accuracy of the existing
Table 2: Algorithm AKLT2 for finding the optimum search direction via a codebook search. \( \Delta(\cdot) \) is a quantizer.

Signal subspace estimates can be poor, leading to inaccurate estimates of the signal. It is therefore desirable to have a mechanism which adjusts the signal subspace dimension to accommodate sudden changes in the signal subspace. One possibility is to estimate the entire set of eigenvectors. This eliminates the need for finding a good search direction since these search spaces are the entire \( N \)-dimensional Euclidean vector space. An additional benefit of having the entire space as the search space is that the estimated eigenvectors computed by the algorithm correspond exactly to those of the sample covariance matrix \( \hat{R}_n \). To find the signal subspace dimension, the transmitter can measure the normalized mean-squared error-like quantity

\[ \rho \equiv \frac{\|x_n - \hat{x}_n\|^2}{\|x_n\|^2} \]

The signal subspace dimension can then be increased until \( \rho \) is below a preset threshold, \( MSE_{\text{max}} \). The transmitter then sends the receiver the KLT coefficients as well as side information consisting of the number of KLT coefficients being transmitted \( (r_{opt}) \) and the number of bits per coefficient. The complete algorithm (AKLT3) is listed in Table 3.
\( \hat{Q}_0 = I_N \)

for \( n = 1, 2, \ldots \)

\[
\begin{align*}
\hat{y}_n &= \hat{Q}_{n-1}^T x_n \\
\hat{y}_n &= \Delta(y_n) \\
\rho &= 1, k = 1
\end{align*}
\]

while \( \rho > MSE_{max} \)

\[
\begin{align*}
\hat{x}_n &= \hat{Q}_{n-1}(:,1:k)\hat{y}_n(1:k); \\
\rho &= \|\hat{x}_n - x_n\|^2/\|x_n\|^2 \\
k &= k + 1;
\end{align*}
\]

if \( k = N + 1 \) AND \( \rho > MSE_{max} \)

Orthonormalize columns of \( \hat{Q}_n \)

\[
k = 1
\]

\[
\begin{align*}
y_n &= \hat{Q}_{n-1}^T x_n; \\
\hat{y}_n &= \Delta(y_n)
\end{align*}
\]

end

end

\( r_{opt} = k - 1 \)

\[
\hat{y}_n(r_{opt} + 1:N) = 0
\]

transmit \( \hat{y}_n, r_{opt} \) and side information to receiver

\[
F = \gamma \hat{\Lambda}_{n-1} + \hat{y}_n\hat{y}_n^T
\]

solve \( F\hat{Q}_n = Q_n\hat{\Lambda}_n \)

end

\( \hat{Q}_0 = I_N \)

for \( n = 1, 2, \ldots \)

wait for \( \hat{y}_n, r_{opt} \) and side information

\[
\begin{align*}
\hat{x}_n &= Q_{n-1}\hat{y}_n \\
F &= \gamma \Lambda_{n-1} + \hat{y}_n\hat{y}_n^T \\
solve F\hat{Q}_n = Q_n\Lambda_n
\end{align*}
\]

end

Table 3: Algorithm AKLT3. This algorithm computes the entire set of eigenvectors and hence does not require a search direction.
6 Experiments

The following signal was generated

\[ x(n) = \begin{cases} 
\cos(0.35\pi n) + \cos(0.78\pi n + 0.35\pi), & n = 1, \ldots, 999 \\
\cos(0.6\pi n) + \cos(0.8\pi n + 0.35\pi), & n = 1000, \ldots, 2,000 
\end{cases} \]

(26)

and algorithm AKLT1 algorithm was applied with \( r = 4 \) and \( \gamma = 0.8 \). The search direction \( v_n \) was set to a zero-mean Gaussian white noise vector. To measure the algorithm’s performance the a priori mean squared error between the original and reconstructed data frame was estimated as

\[ \epsilon_o(n) = \| x_n - \hat{Q}_{n-1} \hat{Q}_n^T x_n \|^2 \]

(27)

where \( \| \| \) is the standard vector 2-norm. We note that this is a more valid measure than a posteriori error since in a compression scenario, the quantity being transmitted is \( \hat{Q}_{n-1} x_n \).

To verify that the KLT basis vectors were indeed being tracked, eigenvectors of the sample correlation matrix were computed and placed in the matrix \( Q_o \). The following measure of the distance between the subspace spanned by the estimated KLT basis vectors in \( \hat{Q}_n \) and the eigenvectors in \( Q_o \) is given by [6]

\[ \xi(n) = \| P_{Q_o} - P_{\hat{Q}_n} \|_F \]

(28)

where \( \| \|_F \) denotes the matrix Frobenius norm and \( P_{Q_o} \) and \( P_{\hat{Q}_n} \) are projection matrices onto the column spaces of \( Q_o \) and \( \hat{Q}_n \), respectively. Figure 1 shows \( \epsilon_o(n) \) for different values of \( N \). As predicted, the convergence speed increases with decreasing \( N \). Figure 2 shows \( \xi(n) \) which demonstrates that the algorithm can indeed track the KLT basis vectors.

The experiment was then repeated using algorithm AKLT2 using a 1000-word Gaussian white noise codebook to find the best search direction as described above. The resulting a priori mean squared error plots are shown in Figure 3. The corresponding subspace error
is shown in Figure 4. The codebook search gives a dramatic improvement in convergence speed.

The experiment was then repeated using algorithm AKLT3, using $MSE_{max} = 10^{-5}$. Since algorithm AKLT3 tracks the eigenvectors of $\hat{R}_n$ exactly, and since the mean squared error between $\hat{x}_n$ and $x_n$ is always minimized by adjusting the signal subspace dimension, we chose as a measure of performance the lowest signal subspace dimension, $r_{opt}$ for which the normalized mean squared error $\rho$ for any given signal frame did not exceed $MSE_{min}$. The resulting plots for $N = 16, 32, \text{ and } 64$ are shown in Fig. 5. It can be seen that since the signal consists of two sinusoids, the signal subspace dimension is 4, consequently, the algorithm has converged when $r_{opt} = 4$. Algorithm AKLT3 was then used in the same experiment but with each KLT coefficient quantized to 8 bits of resolution using a uniform mid-tread quantizer. The maximum allowable normalized mean squared error was set to $MSE_{max} = 0.05$. The corresponding plots of $r_{opt}$ are shown in Figure 6. The higher value of $MSE_{max}$ for the quantizing experiments was necessary in order to meet the error requirement during the step change in signal frequencies at $n = 1000$. Since the KLT coefficients are quantized, a zero mean squared error may not occur even when the signal subspace dimension is $N$. We found that when the KLT coefficients are quantized, the normalized mean squared error ($\|x_n - \hat{x}_n\|^2/\|x_n\|^2$), tends to remain close to, (though never exceeding) $MSE_{max}$. If $MSE_{max}$ is made too small, the specification can only be met for signal subspace dimensions exceeding the theoretical dimension. Though we used a fixed value of $MSE_{max}$ it is possible to lower this threshold during periods of relative stationarity, the larger value of $MSE_{max}$ is only needed during very sudden changes in signal subspace eigenstructure. Figure 7 shows an example of the actual signal frame $x_n$ and the reconstructed signal frame $\hat{x}_n$ obtained by AKLT3 during the quantization experiment.
7 Summary and Discussion

Several algorithms for blindly tracking the KLT basis vectors using only the KLT coefficients and a minimal amount of side information were described. Convergence of the KLT basis vector tracker was proven for a Gaussian white noise search direction. It was shown that for a Gaussian white noise search direction, the convergence is dependent upon the quantity \( N - r \) where \( N \) is the data frame length and \( r \) is the number of KLT basis vectors. The smaller \( N - r \), the faster the convergence. A codebook search method was described which greatly improves the convergence rate of the algorithm. Finally a third algorithm was described which tracks the entire set of eigenvectors and allows for a variable signal subspace dimension.

References


Figure 1: Mean squared error for sinusoid reconstruction for different frame lengths $N$ using algorithm AKLT1 and a white noise search direction.
Figure 2: Subspace tracking error for different frame lengths $N$ using algorithm AKLT1 and a white noise search direction.
Figure 3: Mean squared error for sinusoid reconstruction using a codebook search to find the optimum search direction (algorithm AKLT2). Codebook searching leads to a significant increase in convergence rate.
Figure 4: Subspace tracking error using a codebook search to find the optimum search direction (algorithm AKLT2). Codebook searching leads to a significant increase in convergence rate.
Figure 5: Signal subspace dimension required by algorithm AKLT3 to reduce the normalized mean squared error below $MSE_{max} = 10^{-5}$. 
Figure 6: Signal subspace dimension required by AKLT3 to reduce the normalized mean squared error below $MSE_{max} = 0.05$ using quantized KLT coefficients.
Figure 7: Actual and reconstructed signal frames obtained using quantized KLT coefficients for algorithm AKLT3. The normalized mean squared error for this particular frame was $\rho = 0.0294$. 