DISTRIBUTED KARHUNEN-LOÈVE TRANSFORM WITH NESTED SUBSPACES

Naveen Goela and Michael Gastpar

University of California, Berkeley
Wireless Foundations Center, Department of EECS
Berkeley, CA 94720-1770, USA

ABSTRACT

A network in which sensors observe a common Gaussian source is analyzed. Using a fixed linear transform, each sensor compresses its high-dimensional observation into a low-dimensional representation. The latter is provided to a central decoder that reconstructs the source according to a mean squared error (MSE) distortion metric. The Distributed Karhunen-Loève Transform (*d*-KLT) has been shown to provide a (locally) optimal linear solution for compression at each sensor. While the *d*-KLT achieves the lowest distortion linear reconstruction known, it does not maintain a nested subspace structure. In the case of ideal links to the decoder, this paper presents transforms that maintain nested subspaces, allowing the decoder to approximate a delay-limited source in an online fashion according to a desired sensor schedule. A distortion envelope for one distributed transform with nested subspace properties (*d-nested*-KLT) is provided. In the case of $i.i.d.$ noise to the decoder, under assumptions of power allocation over subspaces, it is also possible to achieve nested subspaces utilizing correlations between sensors' observations. Results are applicable for data access over networks, and online information processing in sensor networks.

Index Terms— Distributed Karhunen-Loève Transform, nested subspaces, distributed compression-estimation.

1. INTRODUCTION

The Karhunen-Loève Transform (KLT) and empirical Principal Components Analysis (PCA) are widely used in fields such as computer vision, biology, signal processing, and information theory. The KLT transform provides a representation of a random source signal using the eigenvectors of the covariance of the source. The representation is useful for source coding, noise filtering, compression, and dimensionality reduction. Eigenfaces for face recognition, sparse PCA for gene analysis, and orthogonal decomposition in transform coding are only a few examples demonstrating the advantage of representing a signal by subspace projections [1] [2] [3].

The *distributed* KLT arises in a network setting illustrated in Fig. (1) in which multiple sensors observe correlated source signals. Effectively, each sensor observes a part of a common source with known covariance structure. The task of designing optimal transforms for each sensor for jointly Gaussian vector sources was defined by Gastpar et. al. and solved via the *d*-KLT iterative algorithm [4] [5]. Simultaneous work by Zhang et. al. for multi-sensor fusion setups also resulted in iterative procedures for sensor transforms [6]. Distributed compression-estimation was extended for non-Gaussian sources, including channel fading and noise effects to model the non-ideal link from sensors to decoder by Schizas et. al. [7]. In current work, Roy and Vetterli [8] provide an *asymptotic* distortion analysis of the *d*-KLT algorithm for two sensors, under certain restrictions that covariance matrices be circulant, in the case when the dimension of the source and observation vectors approaches infinity.

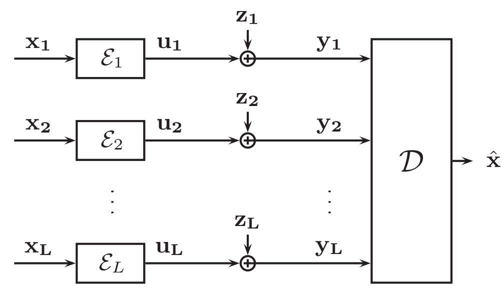


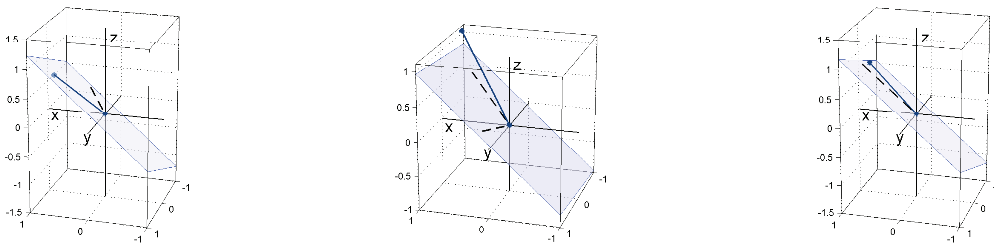
Fig. 1. A network of sensors observes and compresses observations from a central source $\mathbf{x}$. Reduced order descriptions are sent to a decoder.

On a slight tangent, the theory of distributed compressed sensing (DCS) was developed by Baraniuk et. al. [9]. Each sensor computes *random* projections of the data instead of select linear combinations of the data. Assuming a joint sparsity model for signals, DCS theory characterizes how many random projections are necessary for exact reconstruction with low probability of error.

The focus of this paper is to provide a solution for *online* distributed compression-estimation of *delay-limited* Gaussian sources observed by a network of sensors. Unlike standard orthogonal transforms such as the KLT transform, the *d*-KLT distributed transform applied by an encoding sensor is neither orthogonal nor aligned for successive approximation. The *d*-KLT transform is optimized for a fixed number of linear projection measurements per sensor. From the original rate-distortion perspective this is well-motivated; however, in analog-amplitude transmission and in cases where sensors send information incrementally with delay, it is not possible to use the *d*-KLT solution. New distributed transforms exist for online transmission of data. According to a sensor schedule, the decoder receives successive measurements and reconstructs the source incrementally.

The organization of the paper is as follows: Section: (2) presents basic definitions and outlines the problem of nested subspaces; Section: (3) describes the nested subspaces algorithm *d-nested*-KLT for distributed transforms in the ideal case; Section: (4) discusses modifications to the algorithm to account for $i.i.d.$ noise and power constraints. The appendix outlines a closed form derivation for a distributed local transform used iteratively in the nested subspaces algorithm for ideal links.

978-1-4244-2354-5/09/$25.00 ©2009 IEEE



(a) *d-marginal*-KLT (b) *d*-KLT (c) *d-nested*-KLT

Fig. 2. Optimal 1-D and 2-D signal subspaces: Panel 2(a) is the marginal KLT transform (nested+orthogonal) for one sensor. Panel 2(b) is the *d*-KLT transform [4] that is not nested and not orthogonal. Panel 2(c) illustrates the *d-nested*-KLT transform which enforces the nested property. Dark blue solid lines represent best 1-D subspaces. Dotted lines represent additional basis vectors defining 2-D subspaces.

2. PROBLEM STATEMENT

The distributed compression-estimation problem from reduced dimensionality observations is outlined in Fig. (1). Each encoding sensor $\mathcal{E}_\ell$, $\ell = 1 \dots L$, observes a high dimensional Gaussian vector $\mathbf{x}_\ell \in \mathbb{R}^{n_\ell}$ and applies a fixed linear transform $\mathbf{C}_\ell \in \mathbb{R}^{k_\ell \times n_\ell}$ to output a lower dimensional vector $\mathbf{u}_\ell \in \mathbb{R}^{k_\ell}$ where $k_\ell \leq n_\ell$. All observations $\{\mathbf{x}_\ell\}_{\ell=1}^{L}$ are zero mean, jointly Gaussian random vectors and the covariance $\mathbf{\Sigma_x}$ is known where $\mathbf{x} = [\mathbf{x_1}^T\ \mathbf{x_2}^T\ \dots\ \mathbf{x_L}^T]^T$. The task is to design linear transforms $\mathbf{C}_\ell$ for each encoder such that the decoder is able to reconstruct an estimate $\hat{\mathbf{x}}$ of the entire high dimensional source $\mathbf{x}$ with minimal distortion. The decoder receives linear measurements $\mathbf{y}_\ell = \mathbf{C}_\ell \mathbf{x}_\ell + \mathbf{z}_\ell$ in concatenated form:

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_L} \end{bmatrix} = \begin{bmatrix} \mathbf{C_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C_2} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{C_L} \end{bmatrix} \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_L} \end{bmatrix} + \begin{bmatrix} \mathbf{z_1} \\ \mathbf{z_2} \\ \vdots \\ \mathbf{z_L} \end{bmatrix} \tag{1}$$

Equivalently, the task is to design a *block diagonal* transform $\mathbf{C}$ containing individual sensor transforms $\{\mathbf{C}_\ell\}_{\ell=1}^{L}$ on the diagonal:

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{z} \tag{2}$$

For analysis in this paper, the noise is assumed $\mathbf{z} = \mathbf{0}$ in the case of ideal links. In the case of scaled $i.i.d.$ noise, the noise is assumed to be a collection of jointly Gaussian random variables with covariance $\mathbf{\Sigma_z} = \lambda \mathbf{I}$, uncorrelated with all observations $\mathbf{x}$.

For simplicity, it is assumed that all $\mathbf{C}_\ell$ encoder matrices are the same size for compression: $\forall \ell,\ n_\ell = n$ and $\forall \ell,\ k_\ell = k$. In addition, $L = 2$ sensors suffice for explaining distributed algorithms; for $L > 2$, iterative algorithms choose one encoder's transform to optimize, while grouping all other sensor transforms into one virtual fixed transform with block-diagonal structure.

2.1. Distortion Metric

Given measurements $\mathbf{y}$ in Eqn: (2), the optimal MMSE estimate of $\mathbf{x}$ is given by $\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}] = \mathbf{\Sigma_{xy}} \mathbf{\Sigma_y}^{-1} \mathbf{y}$ and the MSE error:

$$D_{MSE} = Tr[\mathbf{\Sigma_x} - \mathbf{\Sigma_x} \mathbf{C}^T (\mathbf{C} \mathbf{\Sigma_x} \mathbf{C}^T + \mathbf{\Sigma_z})^{-1} \mathbf{C} \mathbf{\Sigma_x}] \tag{3}$$

The D_{MSE} trace expression has two useful invariant properties:

Property I: Let $\mathbf{P}_\ell \in \mathbb{R}^{k \times k}$ be any invertible matrix. In the case of ideal links, modifying the transform $\mathbf{C}$ by changing the ℓ^{th} sensor transform as $\mathbf{P}_\ell \mathbf{C}_\ell$ does not alter the D_{MSE}. Thus, only the subspaces spanned by the row measurement vectors of each $\mathbf{C}_\ell$ determine the MSE error. This follows from Eqn: (3) if $\mathbf{z} = \mathbf{0}$.

Property II: Let $\mathbf{P}_\ell \in \mathbb{R}^{k \times k}$ be any *orthogonal* matrix such as a rotation. In the case of scaled $i.i.d.$ noise, modifying the transform $\mathbf{C}$ by changing the ℓ^{th} sensor transform as $\mathbf{P}_\ell \mathbf{C}_\ell$ does not alter the D_{MSE}. This follows from Eqn: (3) if $\mathbf{\Sigma_z} = \lambda \mathbf{I}$.

Along with the D_{MSE} expression, we define a quantity $D_{MSE}[k]$ which is the distortion at the decoder when all sensors compute and send exactly k projections of their length n observations. If noise $\mathbf{z} = \mathbf{0}$, the goal is to choose matrices $\mathbf{C}_\ell \in \mathbb{R}^{k \times n}$ that minimize $D_{MSE}[k]$. In the presence of noise, an additional output power constraint per sensor is required: $Tr[\mathbf{C}_\ell \mathbf{\Sigma}_{\mathbf{x}_\ell} \mathbf{C}_\ell{}^T] \leq P_\ell$, see [7].

2.2. Encoding Strategies

The following encoding matrices $\mathbf{C}_\ell$ achieve different distortion results $D_{MSE}[k]$ at the decoder:

- Joint KLT (***joint*-KLT**): The joint KLT sets $\mathbf{C} = [\mathbf{Q_x}^{(kL)}]^T$, the top (kL) eigenvectors corresponding to the largest eigenvalues in the SVD decomposition of covariance $\mathbf{\Sigma_x}$. This is the standard KLT solution, but not a distributed solution since $\mathbf{C}$ is not block-diagonal. There is effectively only one sensor with full access to observations $\mathbf{x}$.

- Marginal KLT (***d-marginal*-KLT**): The marginal KLT applies the KLT transform to each sensor, ignoring inter-sensor correlations. Thus, $\mathbf{C}_\ell = [\mathbf{Q}_{\mathbf{x}_\ell}{}^{(k)}]^T$, the matrix of k top eigenvectors in the SVD decomposition of covariance $\mathbf{\Sigma}_{\mathbf{x}_\ell}$. If noise is $i.i.d.$, the optimal subspaces are similar, but each basis vector is scaled according to power allocation over the subspaces.

- Distributed KLT (***d*-KLT**): In the noiseless setting, the distributed KLT is an iterative algorithm, see [4], that optimizes the encoding matrix $\mathbf{C}_\ell$ of one sensor ℓ, fixing all other encoding matrices. This algorithm takes into account the correlations between sensor observations. In the case of noise, see [7] for details, a more complex distributed algorithm is possible with power constraints handled iteratively as well.

2.3. Nested Subspaces Problem

The above encoding strategies achieve a value $D_{MSE}[k]$ for fixed k projections per sensor. The *joint*-KLT centralized encoding is globally optimal for all k. The *d-marginal*-KLT encoding is sub-optimal, yet is distributed and accommodates delay-limited Gaussian sources, because subspaces spanned by the row vectors of each $\mathbf{C}_\ell$ transform are orthogonal. The solution optimizing $D_{MSE}[k]$ also includes the row vectors for the solution that optimizes $D_{MSE}[k-1]$. We define this as the *nested subspaces property*, illustrated in Fig. (2). Lastly, the *d*-KLT encoding does *not* accommodate delay-limited Gaussian sources. The subspaces defined by the row basis vectors of $\mathbf{C}_\ell$ are not orthogonal or nested, but do take into account inter-sensor signal correlations.

The nested subspaces problem is to design matrices $\mathbf{C}_\ell$ that achieve low MSE distortion simultaneously for $D_{MSE}[i]$, $i = 1, 2, \ \ldots \ k$. We define a nested subspace configuration for matrices $\mathbf{C}_\ell$ to be strictly superior to another configuration if and only if $D_{MSE}[i]$ is less $\forall i$. For this paper, we assume a symmetric sensor schedule: each sensor sends one component at a time. The decoder receives L measurements at a time for incremental reconstruction.

One solution for nested subspaces is to use the *d*-KLT solution directly. The *d*-KLT solution ensures $D_{MSE}[k]$ is optimized for exactly k projections; however, earlier subspaces with distortions $D_{MSE}[i]$, $i < k$ are not guaranteed to be low distortion subspaces and may incur higher distortion than the *d-marginal*-KLT subspaces. A second solution is to select $D_{MSE}[1]$, and choose subsequent subspaces conditioned on the fact that the first measurements were sent. But, this strategy also leads to subspace configurations that are not strictly superior to the *d-marginal*-KLT solution (committing early to subspaces restricts future aligned subspaces). This paper constructs new transforms for online reconstruction that are strictly superior to the *d-marginal*-KLT transforms.

3. NESTED SUBSPACES ALGORITHM: IDEAL LINKS

The distributed nested KLT (*d-nested*-KLT) aims to align subspaces in a similar way to basic orthogonal transforms such as the KLT. Encoding matrices $\mathbf{C}_\ell$ are initialized to the *d-marginal*-KLT encoding matrices, which have the nested subspaces property. Next, the *d-nested*-KLT algorithm exploits inter-sensor correlations by iteratively optimizing each $\mathbf{C}_\ell$ such that $D_{MSE}[i]$ decreases but $D_{MSE}[i-1]$ and $D_{MSE}[i+1]$ remain constant, and subspaces remain nested. To optimize a selected $\mathbf{C}_\ell$, an iterative algorithm fixes all other matrices, and finds the best local transform. To maintain nested subspaces, a modified local transform is required.

3.1. Local Nested Transform

In the ideal case $\mathbf{z} = 0$, power allocation is unnecessary, and adjusting subspaces is possible via **Property I** of the D_{MSE} error. Assume $L = 2$ sensors make observations $\mathbf{x} = [\mathbf{x_1}^T \ \mathbf{x_2}^T]^T$ where $\mathbf{x_1} \in \mathbb{R}^n, \mathbf{x_2} \in \mathbb{R}^n$ with known covariance:

$$\mathbf{\Sigma_x} = \begin{bmatrix} \mathbf{\Sigma_1} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma_2} \end{bmatrix} \tag{4}$$

The task at the decoder is to reconstruct source $\mathbf{x}$ using observations $\mathbf{y_1} = \mathbf{B_1 S_1 x_1}, \mathbf{f_1} = \mathbf{F_1 x_1}, \mathbf{y_2} = \mathbf{C_2 x_2}$.

Matrices $\mathbf{S_1} \in \mathbb{R}^{s_1 \times n}$, $\mathbf{F_1} \in \mathbb{R}^{f_1 \times n}$, and $\mathbf{C_2} \in \mathbb{R}^{k \times n}$ represent fixed transforms. $\mathbf{F_1}$ contains pre-selected subspace directions for the first sensor that have already been sent to the decoder. $\mathbf{C_2}$ represents the second sensor held fixed. The special matrix $\mathbf{S_1}$ contains the subspace over which we would like to search for new optimal directions. The new directions *in the subspace of* $\mathbf{S_1}$ are obtained by multiplying on the left by a matrix $\mathbf{B_1} \in \mathbb{R}^{b_1 \times s_1}$ that is to be optimized. The optimization problem is to find the best transform $\mathbf{B_1}$ for the first sensor that will yield the best LMMSE estimate of source $\mathbf{x}$ at the decoder. This is given via the following theorem:

Theorem 1 *Given observations* $\mathbf{y_1} = \mathbf{B_1 S_1 x_1}$, $\mathbf{f_1} = \mathbf{F_1 x_1}$, *and* $\mathbf{y_2} = \mathbf{C_2 x_2}$, *for a fixed* $s_1 \times n$ *matrix* $\mathbf{S_1}$, *a fixed* $f_1 \times n$ *matrix* $\mathbf{F_1}$, *and a fixed* $k \times n$ *matrix* $\mathbf{C_2}$, *assuming* $\mathbf{z} = \mathbf{0}$, *LMMSE reconstruction of the jointly Gaussian random vector* $\mathbf{x}$ *attains the smallest MSE if* $\mathbf{B_1} \in \mathbb{R}^{b_1 \times s_1}$ *is chosen as* ($b_1 < s_1 < n$, $f_1 < n$):

$$\mathbf{B_1} = (\mathbf{Q_w}^{(b_1)})^T \begin{bmatrix} \mathbf{J_1} \\ \mathbf{A_1} \end{bmatrix} \tag{5}$$

where $\mathbf{Q_w}$, $\mathbf{J_1}$, *and* $\mathbf{A_1}$ *are defined in the Appendix.*

3.2. Nested Subspaces Algorithm

Using the above local transform in the ideal case, the *d-nested*-KLT algorithm iteratively decreases $D_{MSE}[i]$ for $i = 1, 2, \ \ldots \ k$, maintaining nested subspaces. This algorithm is described by Algorithm: (1) for $L = 2$ sensors without loss of generality.

Algorithm 1 *d-nested*-KLT

1: Assume, $L = 2$, $k_\ell = k$, $n_\ell = n$, $\mathbf{\Sigma_x}$ is known, $\mathbf{z} = 0$. Initialize $\mathbf{C}_\ell$ for $\ell = 1, 2$ to the *d-marginal*-KLT distributed solution: $\mathbf{C_1} = \left[\mathbf{Q_{x_1}^{(k)}}\right]^T$ and $\mathbf{C_2} = \left[\mathbf{Q_{x_2}^{(k)}}\right]^T$.
2: **for** $i = 1$ to $k - 1$ **do**
3:   Update i^{th} row basis vector for $\mathbf{C_1}$ and $\mathbf{C_2}$:
4:   **while** $D_{MSE}[i]$ decreases **do**
5:     Select a random $\mathbf{C}_\ell$ to optimize, assume we select $\mathbf{C_1}$. Hold $\mathbf{C_2}$ fixed.
6:     Let $\mathbf{F_1}$ be the matrix comprised of the first $(i-1)$ row vectors of $\mathbf{C_1}$. If $i = 1$, $\mathbf{F_1} = [\,]$, an empty matrix.
7:     Let $\mathbf{S_1}$ be the matrix comprised of the i^{th} and $(i+1)^{th}$ row vectors of $\mathbf{C_1}$.
8:     Compute the optimal $\mathbf{B_1}$ matrix ($b_1 = 1$) using the modified local transform, see Theorem 1.
9:     Modify $\mathbf{C_1}$ by setting its i^{th} row to be $\mathbf{B_1 S_1}$.
10:     Ensure that the subspace spanned by the i^{th} and $(i+1)^{th}$ row vectors of $\mathbf{C_1}$ has not changed.
11:   **end while**
12: **end for**
13: GOTO *for* loop in Step 2 to decrease D_{MSE} further if possible.

The results for the nested algorithm are shown in Fig. (3) for a random p.s.d. covariance matrix $\mathbf{\Sigma_x}$ with $n = 6$ and $L = 2$. The *d*-KLT solution achieves the lowest $D_{MSE}[k]$ for a fixed k; however, the *d-nested*-KLT solution applies to delay-limited sources that are reconstructed incrementally. By construction, the *d-nested*-KLT solution is strictly superior to the *d-marginal*-KLT solution, because it exploits inter-sensor correlations.

4. NOISE CONSIDERATIONS

In a realistic setting, the presence of noise forces a power constraint per sensor: $Tr[\mathbf{C}_\ell \mathbf{\Sigma}_{\mathbf{x}_\ell} \mathbf{C}_\ell^T] \leq P_\ell$, see Schizas et. al. [7] for distributed compression-estimation with power constraints. This section provides a motivation for nested subspaces in the presence of additive white noise.

In the high SNR regime, the *d-nested*-KLT for ideal links will achieve a lower $D_{MSE}[i]$ for all i than the *d-marginal*-KLT solution. In the low SNR regime, aligning subspaces becomes a constrained optimization problem due to power allocation. Assuming a fixed power allocation, the arrangement of subspaces alone affects distortion. The following iterative procedure for noisy links is similar to the ideal link case. First, matrices $\mathbf{C}_\ell$ are initialized to the *d-marginal*-KLT solution with optimal power allocation. In each iteration, the trace error expression (Eqn: (3)) with $i.i.d.$ noise must be minimized iteratively as in Algorithm: (1). In the ideal case, a closed form formula for the local transform was used, exploiting **Property I**. For $i.i.d.$ noise, given a fixed power allocation (scaling) of each basis vector in encoding matrices $\mathbf{C}_\ell$, it is possible to adjust basis vectors to exploit inter-sensor correlations (as in Step 8-9 of Algorithm (1)) by multiplying by an *orthogonal* matrix to lower $D_{MSE}[i]$. Lowering $D_{MSE}[i]$ but leaving $D_{MSE}[i+1]$ unchanged is possible via **Property II** of the trace expression. The power allocation over subspaces is unaffected, but $D_{MSE}[i]$ for $i < k$ is reduced iteratively (as in the ideal case) and subspaces remain nested.

5. CONCLUSION

We have shown the existence of encoding matrices for online distributed compression-estimation that are strictly superior to matrices that ignore inter-sensor correlations. The problem of distributed online reconstruction is challenging because all subspaces must be aligned to send successive components. In some sense, this concept parallels successive refinement of sources in rate-distortion theory. By understanding nested subspaces, we gained understanding of the arrangement of (locally) optimal subspaces in the distributed case. For future work, it is interesting to see whether there are any other nested subspace configurations that are strictly superior to the ones initialized from the *d-marginal*-KLT subspaces in this paper.

6. APPENDIX

The derivation of Theorem 1 follows from the proof of the original *d*-KLT local transform proof, see details given by Gastpar et. al. [4].

$\mathbf{J_1} = \mathbf{J}^{(\mathbf{s_1})}$ is the first s_1 columns of the matrix $\mathbf{J}$, and $\mathbf{A_1} = \mathbf{A}^{(\mathbf{s_1})}$ is the first s_1 columns of the matrix $\mathbf{A}$ computed as:

$$\mathbf{J} = \begin{bmatrix} \Sigma_1 \mathbf{S_1}^T & \Sigma_1 \mathbf{F_1}^T & \Sigma_{12} \mathbf{C_2}^T \end{bmatrix} \Phi^\dagger \tag{6}$$

$$\mathbf{A} = \begin{bmatrix} \Sigma_{21} \mathbf{S_1}^T & \Sigma_{21} \mathbf{F_1}^T & \Sigma_2 \mathbf{C_2}^T \end{bmatrix} \Phi^\dagger \tag{7}$$

where $\mathbf{\Phi}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{S_1}\Sigma_1\mathbf{S_1}^T & \mathbf{S_1}\Sigma_1\mathbf{F_1}^T & \mathbf{S_1}\Sigma_{12}\mathbf{C_2}^T \\ \mathbf{F_1}\Sigma_1\mathbf{S_1}^T & \mathbf{F_1}\Sigma_1\mathbf{F_1}^T & \mathbf{F_1}\Sigma_{12}\mathbf{C_2}^T \\ \mathbf{C_2}\Sigma_{21}\mathbf{S_1}^T & \mathbf{C_2}\Sigma_{21}\mathbf{F_1}^T & \mathbf{C_2}\Sigma_2\mathbf{C_2}^T \end{bmatrix} \tag{8}$$

Under the following constraint:

$$Rank \begin{bmatrix} \mathbf{S_1} \\ \mathbf{F_1} \end{bmatrix} = s_1 + f_1 \leq n \tag{9}$$

the symbol $\mathbf{\Phi}^\dagger$ is equivalent to a matrix inverse. More generally, an infinite number of solutions may occur if $\mathbf{S_1}$ and $\mathbf{F_1}$ are redundant; in this case, the † implies the Moore-Penrose pseudo-inverse.

The matrix $\mathbf{Q_w}^{(b_1)}$ is the first b_1 columns of the matrix $\mathbf{Q_w}$ obtained from the eigendecomposition of covariance $\mathbf{\Sigma_w}$ defined:

$$\mathbf{\Sigma_w} = \mathbf{Q_w}\mathbf{\Lambda_w}\mathbf{Q_w}^T = \begin{bmatrix} \mathbf{J_1S_1} \\ \mathbf{A_1S_1} \end{bmatrix} \mathbf{\Psi} \begin{bmatrix} \mathbf{J_1S_1} \\ \mathbf{A_1S_1} \end{bmatrix}^T \tag{10}$$

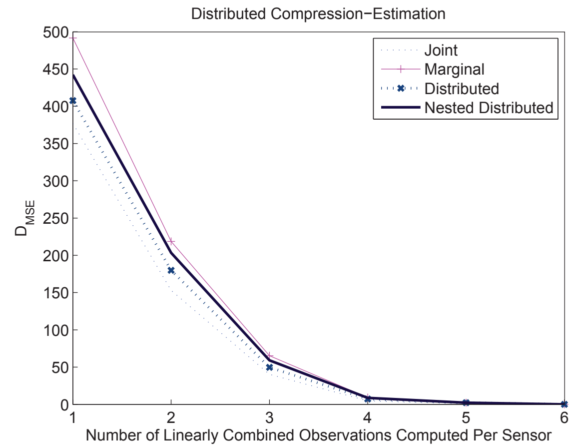


Fig. 3. The *d-nested*-KLT solution maintains nested subspaces for incremental reconstruction at the decoder. The *d*-KLT solution is optimized for a fixed k linearly combined measurements per sensor, but is not applicable for delay-limited Gaussian sources.

$$\mathbf{\Psi} = \mathbf{\Sigma_1} - \begin{bmatrix} \mathbf{\Sigma_1F_1^T} & \mathbf{\Sigma_{12}C_2^T} \end{bmatrix} \mathbf{\Upsilon}^{-1} \begin{bmatrix} \mathbf{F_1\Sigma_1} \\ \mathbf{C_2\Sigma_{21}} \end{bmatrix} \tag{11}$$

$$\mathbf{\Upsilon} = \begin{bmatrix} \mathbf{F_1\Sigma_1F_1^T} & \mathbf{F_1\Sigma_{12}C_2^T} \\ \mathbf{C_2\Sigma_{21}F_1^T} & \mathbf{C_2\Sigma_2C_2^T} \end{bmatrix} \tag{12}$$

7. REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] A. d'Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. Lanckriet, "A direct formulation for sparse pca using semidefinite programming," *SIAM Review*, no. 49, pp. 434–448, 2007.

[3] V. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, no. 5, 2001.

[4] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed karhunen loève transform," *IEEE Transactions on Information Theory*, vol. 52, pp. 5177–5196, 2006.

[5] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed karhunen-loève transform," in *Proc. IEEE Workshop on Multimedia Signal Processing*, (St. Thomas, Virgin Islands, USA), pp. 57–60, Dec. 2002.

[6] K. Zhang, *Best Linear Unbiased Estimation Fusion with Constraints*. PhD thesis, Univ. of New Orleans, New Orleans, LA, 2003.

[7] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, pp. 4284–4299, August 2007.

[8] O. Roy and M. Vetterli, "Dimensionality reduction for distributed estimation in the infinite dimensional regime," *IEEE Transactions on Information Theory*, vol. 54, pp. 1655–1669, April 2008.

[9] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Available at: http://dsp.rice.edu/cs.