A GENERALIZED OPTIMAL CORRELATING TRANSFORM FOR MULTIPLE DESCRIPTION CODING AND ITS THEORETICAL ANALYSIS

Stefan Uhlich and Bin Yang

Chair of System Theory and Signal Processing
University of Stuttgart, Germany

ABSTRACT

This paper considers a coding scheme for data transmission over erasure channels which is also known as multiple description coding. The LMMSE prefilter method of Romano [1] is reviewed and generalized to allow three different operational modes of the prefilter. They include the possibility to decrease or increase the number of descriptions to be transmitted. We derive explicitly the Hessian matrix for an efficient calculation of the prefilter. We also study the properties of the distortion measure theoretically.

***Index Terms*—** multiple description coding, transform coding, correlating transform

1. INTRODUCTION

Multiple description coding (MDC) is often linked with a packet oriented transmission scheme like the internet. In the internet, some packets (i.e. descriptions) might get lost. This may e.g. be the case for an internet-router that is congested and its buffers overflow. The problem at the receiver is now to obtain an estimate of the original information from the subset of available packets.

But multiple description coding also appears to be a valid tool for an incremental specification of signals. In the Collaborative Research Center SFB 732 [2], methods for incremental specification of speech are investigated. One feature that one would expect from such an incremental scheme is that subsets of different descriptions of the speech signal can be arbitrarily chosen and help to restore a better representation of the original speech sample.

A good overview of current techniques for multiple description coding can be found in [3, 4].

In this paper we investigate a correlating transform which was introduced in the inspiring paper [1]. After a short introduction into the MDC problem, we will further generalize [1] in section 3. This generalization allows us to handle the case that more descriptions are used after the correlating transform. Thus, our scheme provides the possibility of a redundancy coding. This new operational mode of a MDC correlating transform is very interesting in the case that the channel offers enough bandwidth. It allows us to further minimize the distortion by transmitting redundant descriptions. At the end of section 3, we summarize all possible operational modes. Section 4 finally is a collection of characteristics that the distortion measure inhibits. Its Hessian matrix is derived which is useful for solving the optimization problem. A Matlab Toolbox with an implementation of our proposed algorithm can be found in [5].

Following notation is used throughout this paper: $\underline{x}$ denotes a vector, $\mathbf{X}$ a matrix and $\mathbf{I}$ the identity matrix.

This work is supported in part by the Deutsche Forschungsgemeinschaft (Collaborative Research Center SFB 732: Incremental specification in context)

2. THE ERASURE CHANNEL AND THE OPTIMAL CORRELATING TRANSFORM

Fig. 1 shows the considered system. The real-valued, zero-mean random vector $\underline{x}$ is to be transmitted over an erasure channel that might randomly erase elements of $\underline{x}$ and therefore the received vector $\underline{z}$ might be of smaller dimension than $\underline{x}$. Note, that we do not consider quantization noise as was done in [1]. We will assume in this paper that unerased descriptions are received without distortion. The erasure process of the channel is described by a matrix $\mathbf{P}_e$ which is composed of the unit row vectors of the surviving descriptions, i.e. it has zero columns for the descriptions that do not survive. A further assumption is that $\mathbf{P}_e$ is known to the receiver, i.e. it knows which descriptions got lost during transmission.

In the case that there is no precoding (i.e. $\mathbf{T} = \mathbf{I}$), the linear MMSE estimation of $\underline{x}$ from the received vector $\underline{z}$ is clearly (see e.g. [6])

$$\hat{\underline{x}} = \mathbf{R}\mathbf{P}_e^T(\mathbf{P}_e\mathbf{R}\mathbf{P}_e^T)^{-1}\underline{z} \tag{1}$$

where we used $\mathbf{R} = \mathrm{E}[\underline{x}\,\underline{x}^T]$. The corresponding correlation matrix of the error $\underline{\epsilon} = \underline{x} - \hat{\underline{x}}$ is

$$\mathbf{R}_{\underline{\epsilon}} = \mathbf{R} - \mathbf{R}\mathbf{P}_e^T(\mathbf{P}_e\mathbf{R}\mathbf{P}_e^T)^{-1}\mathbf{P}_e\mathbf{R} \tag{2}$$

and we can define a distortion D_e as

$$D_e = \mathrm{E}\left[\|\underline{\epsilon}\|^2\right] = \mathrm{tr}\left\{\mathbf{R}_{\underline{\epsilon}}\right\} \tag{3}$$

Eq. (1) allows us to estimate the lost descriptions of $\underline{x}$. However, this is only a reaction of the receiver to a particular erasing matrix $\mathbf{P}_e$. It is obvious that we can decrease (3) if we

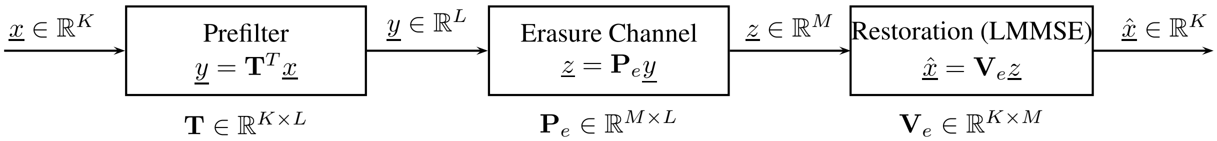

Fig. 1. Erasure Channel

use a precoding. In [1], Romano dealt with this problem. He introduced a transform $\mathbf{T}$ before the data vector $\underline{x}$ is transmitted over the erasure channel. The idea of this transform is to distribute the *important* information of $\underline{x}$ over all elements of $\underline{y}$. Since different erasure constellations $\mathbf{P}_e$ are possible, $\mathbf{T}$ has to be designed to consider all of them.

The encoding and decoding equations with a prefilter $\mathbf{T} \in \mathbb{R}^{K \times L}$ and $L \leq K$ are given by

$$\underline{y} = \mathbf{T}^T \underline{x} \tag{4}$$

$$\hat{\underline{x}} = \mathbf{V}_e \underline{z} = \mathbf{R}\mathbf{T}\mathbf{P}_e^T \left(\mathbf{P}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\mathbf{P}_e^T\right)^{-1} \underline{z} \tag{5}$$

The optimal transform $\mathbf{T}$ is found by minimizing the overall distortion D which takes into account all possible erasure constellations $\mathbf{P}_e$ by calculating a weighted sum of the individual distortions D_e.

$$D = \sum_{e=1}^{E} w_e D_e = \sum_{e=1}^{E} w_e \Big[\operatorname{tr}(\mathbf{R}) - \operatorname{tr}\Big(\mathbf{R}\mathbf{T}\mathbf{P}_e^T \left(\mathbf{P}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\mathbf{P}_e^T\right)^{-1} \mathbf{P}_e \mathbf{T}^T \mathbf{R} \Big) \Big]. \tag{6}$$

E is the total number of error constellations and w_e the weighting of a particular error constellation. w_e can e.g. be chosen to be the probability that the error constellation $\mathbf{P}_e$ occurs. As a simple model, we could e.g. assume that each description is safely transmitted with the probability p. Therefore, the probability that one specific error constellation $\mathbf{P}_e$ with $L - M$ erased descriptions occurs is merely $w_e = p^M (1-p)^{L-M}$ and the total number of error constellations is $E = 2^L$.

If $E = 1$, i.e. only one $\mathbf{P}_e$ is possible, then we can easily minimize D by using only the descriptions that are not erased. These descriptions have to contain the coordinates of $\underline{x}$ along the eigenvectors of $\mathbf{R}$ with the largest eigenvalues. This corresponds to the Karhunen-Loeve transform which is the optimal linear transform $\mathbf{T}$ in this case.

For $E > 1$, no closed-form solution is known and Romano proposed a gradient search to seek for the optimal transform $\mathbf{T}$ in [1].

3. A GENERALIZED CORRELATING TRANSFORM

In this section, we will generalize the precoding in (4). This generalization will allow us to also have more descriptions after the precoding, i.e. $L > K$. The basic idea is to transmit redundant descriptions which are more unlikely to be erased altogether and therefore more descriptions remain available to compute a better estimate $\hat{\underline{x}}$ at the expense of an increasing bandwidth. The only difficulty is a possible rank deficient correlation matrix of $\underline{z}$ if the redundancy introduced by the transform $\mathbf{T}$ is not completely erased by the channel.

As an example, let $K = 2$ and $L = 4$, i.e. the prefilter adds two redundant descriptions. Assume further that the channel transmits all descriptions, i.e. $\mathbf{P}_e = \mathbf{I}$. The problem now is that $\mathbf{P}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\mathbf{P}_e^T = \mathbf{T}^T \mathbf{R}\mathbf{T} \in \mathbb{R}^{4 \times 4}$ has at most a rank of two and can therefore not be inverted.

We propose the following selection process to solve this problem: As we seek for the optimal $\mathbf{T}$ by an iterative method like gradient search in [1] or Newton method, we determine at each iteration the rank of $\mathbf{T}\mathbf{P}_e^T$ which corresponds to the number of non-redundant descriptions. If $\operatorname{rank}\{\mathbf{T}\mathbf{P}_e^T\} < M$, i.e. $\mathrm{E}[\underline{z}\,\underline{z}^T]$ will be rank deficient, we delete additional rows of $\mathbf{P}_e$ resulting in $\tilde{\mathbf{P}}_e \in \mathbb{R}^{\tilde{M} \times L}$ ($\tilde{M} \leq M$) such that $\operatorname{rank}\{\mathbf{T}\tilde{\mathbf{P}}_e^T\} = \tilde{M}$. This corresponds to deleting redundant descriptions in $\underline{z}$ such that the vector of non-redundant transmitted descriptions has a full rank correlation matrix. The question is which rows of $\mathbf{P}_e$ should be deleted. We systematically try all possible combinations of row vectors of $\mathbf{P}_e$ and use that combination with the smallest condition number of $\mathbf{T}\tilde{\mathbf{P}}_e$. This also ensures that the calculation is numerically robust. The generalized distortion function therefore is

$$D = \sum_{e=1}^{E} w_e \Big[\operatorname{tr}(\mathbf{R}) - \operatorname{tr}\Big(\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T \left(\tilde{\mathbf{P}}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1} \tilde{\mathbf{P}}_e \mathbf{T}^T \mathbf{R} \Big) \Big]. \tag{7}$$

$\tilde{\mathbf{P}}_e$ includes both, the erasure process of the channel and the selection process above to handle redundant descriptions.

Note, that if we had not neglected the influence of the noise (e.g. because of quantization noise), then instead of $\mathbf{P}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\mathbf{P}_e^T$ we would have to consider the full rank matrix $\mathbf{P}_e \mathbf{T}^T \mathbf{R}\mathbf{T}\mathbf{P}_e^T + \mathbf{R}_{\text{noise}}$. In this case, the above procedure is not necessary and all descriptions will help to improve the estimation of the original $\underline{x}$.

In total, we can now distinguish between three different operational modes of the correlating transform $\mathbf{T} \in \mathbb{R}^{K \times L}$.

$K > L$: The vector $\underline{y}$ that is transmitted has a smaller number of elements, i.e. the prefilter $\mathbf{T}$ performs a compression. Additionally, it will try to minimize the distortion D and therefore $\mathbf{T}$ will mainly consist of a combination of the eigenvectors with large eigenvalues.

$K = L$: In this case, $\underline{y}$ and $\underline{x}$ have the same dimension and we improve the robustness of the transmission against the erasure channel by the precoding $\mathbf{T}$.

$K < L$: The number of descriptions increases after the precoding. This corresponds to a redundancy coding, where we allow the transmitter to use more descriptions to safely transmit the information to the receiver.

4. PROPERTIES OF THE OPTIMIZATION PROBLEM

Below, we study the generalized correlating transform theoretically. In particular, we give four properties of the distortion D in (7). Especially the knowledge of the Hessian matrix will help us to find the optimal transform $\mathbf{T}$ efficiently.

4.1. Gradient and Hessian of D

Property 1: Eq. (8) and (9) give the elements of the gradient vector and the Hessian matrix of D where $\mathbf{W} = \mathbf{V}_e\tilde{\mathbf{P}}_e$. ∂_α denotes the derivative with respect to α and $\partial_{\alpha\beta}$ is the second derivative with respect to α and β.

Proof: Note, that the gradient vector was already stated in [1]. Because of space limitations, we omit a detailed derivation of (9) in this paper. ■

The knowledge of the Hessian matrix allows us to use more complex optimization methods, e.g. the Newton method, to determine the optimal $\mathbf{T}$ that minimizes (7).

We can simplify (8) and (9) further if α and β denote two elements of $\mathbf{T}$ at the positions (i, j) and (k, l) by using

$$\partial_\alpha \mathbf{T} = \partial_{ij}\mathbf{T} = \mathbf{J}^{ij} \tag{10a}$$

$$\partial_{\alpha\beta}\mathbf{T} = \partial_{ij,kl}\mathbf{T} = \mathbf{0}. \tag{10b}$$

$\mathbf{J}^{ij} \in \mathbb{R}^{K\times L}$ is a single entry matrix which is zero everywhere except for a "1" at the position (i, j).

4.2. Stationary points of D

Property 2: Each transform $\mathbf{T} = \mathbf{U}$ with $L \leq K$, where the columns of $\mathbf{U}$ contain any distinctive eigenvectors of $\mathbf{R} = \mathrm{E}[\underline{x}\,\underline{x}^T]$, is a stationary point of the distortion (7).

Proof: We consider only one error constellation $\tilde{\mathbf{P}}_e$ in (8), i.e. one summand. For the special case that $\mathbf{T} = \mathbf{U}$, we obtain after some calculations

$$\begin{aligned}
\mathbf{W} &= \mathbf{R}\mathbf{U}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{U}^T\mathbf{R}\mathbf{U}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e \\
&= \mathbf{U}\boldsymbol{\Lambda}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\boldsymbol{\Lambda}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e \\
&= \mathbf{U}\tilde{\mathbf{P}}_e^T\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^{-1}\tilde{\mathbf{P}}_e = \mathbf{U}\tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e
\end{aligned} \tag{11}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{R}$ to the eigenvectors $\mathbf{U}$ and $\tilde{\boldsymbol{\Lambda}} = \tilde{\mathbf{P}}_e\boldsymbol{\Lambda}\tilde{\mathbf{P}}_e^T$. In the appendix, we show that $\boldsymbol{\Lambda}\tilde{\mathbf{P}}_e^T = \tilde{\mathbf{P}}_e^T\tilde{\boldsymbol{\Lambda}}$ holds. This was used in the last line of (11). Therefore, (8) can be rewritten as

$$\begin{aligned}
\frac{\partial D}{\partial_\alpha} &= 2\sum_{e=1}^{E} w_e \operatorname{tr}\left\{\tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e\mathbf{U}^T\!\left(\mathbf{U}\tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e\mathbf{U}^T - \mathbf{I}\right)\mathbf{R}\partial_\alpha\mathbf{T}|_{\mathbf{T}=\mathbf{U}}\right\} \\
&= 2\sum_{e=1}^{E} w_e \operatorname{tr}\{\mathbf{0}\} = 0
\end{aligned}$$

as $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $\tilde{\mathbf{P}}_e\tilde{\mathbf{P}}_e^T = \mathbf{I}$. ■

Property 2 is even valid for the case that $\mathbf{T}$ contains eigenvectors of $\mathbf{R}$ several times. The selection process introduced in section 3 will erase the duplicated eigenvectors and $\mathbf{U}\tilde{\mathbf{P}}_e^T$ will only contain distinctive eigenvectors.

4.3. Normalization of $\mathbf{T}$

Property 3: Let $\mathbf{N} = \operatorname{diag}(n_{11}, \ldots, n_{LL}) \in \mathbb{R}^{L\times L}$ be a diagonal matrix and $n_{ii} \neq 0$. Then $\mathbf{TN}$ has the same distortion D as $\mathbf{T}$, i.e. (7) is invariant to a scaling of the columns of $\mathbf{T}$.

Proof: Analog to the proof of property 1, we consider only one term in (7)

$$\begin{aligned}
&\mathbf{R}\mathbf{T}\mathbf{N}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{N}\mathbf{T}^T\mathbf{R}\mathbf{T}\mathbf{N}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e\mathbf{N}\mathbf{T}^T\mathbf{R} \\
&= \mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\tilde{\mathbf{N}}\tilde{\mathbf{N}}^{-1}\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{N}}^{-1}\tilde{\mathbf{N}}\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R} \\
&= \mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}
\end{aligned} \tag{12}$$

where $\tilde{\mathbf{N}} = \tilde{\mathbf{P}}_e\mathbf{N}\tilde{\mathbf{P}}_e^T \in \mathbb{R}^{M\times M}$. Here, we again use the identity $\mathbf{N}\tilde{\mathbf{P}}_e^T = \tilde{\mathbf{P}}_e^T\tilde{\mathbf{N}}$ from the appendix. ■

From the property above we see that each column vector of $\mathbf{T}$ can be normalized to one. We can therefore reduce the computational complexity by constraining $\mathbf{T}$ to a special structure, e.g. setting the norm of each column vector to one. This is done in this work by using spherical coordinates [7]. The jth column of $\mathbf{T}$ using spherical coordinates is

$$\begin{bmatrix}
\cos\phi_{1j} \\
\cos\phi_{2j}\sin\phi_{1j} \\
\vdots \\
\cos\phi_{(K-1)j}\prod_{i=1}^{K-2}\sin\phi_{ij} \\
\prod_{i=1}^{K-1}\sin\phi_{ij}
\end{bmatrix} \tag{13}$$

where $0 \leq \phi_{ij} \leq \pi$ for $i = 1, \ldots, K-2$ and $0 \leq \phi_{(K-1)j} < 2\pi$. By using spherical coordinates, the total number of unknowns is reduced from KL to $(K-1)L$. A drawback of spherical coordinates is that $\partial_\alpha\mathbf{T}$ has more than one non-zero entry and $\partial_{\alpha\beta}\mathbf{T}$ is not always zero in comparison to (10) because α and β now denote the angles ϕ_{ij}. This might be the

$$\frac{\partial D}{\partial_\alpha} = 2\sum_{e=1}^{E} w_e \operatorname{tr}\left\{\mathbf{W}^T\left(\mathbf{W}\mathbf{T}^T - \mathbf{I}\right)\mathbf{R}\partial_\alpha\mathbf{T}\right\} \tag{8}$$

$$\frac{\partial^2 D}{\partial_\alpha\partial_\beta} = 2\sum_{e=1}^{E} w_e \operatorname{tr}\Big\{\Big(\mathbf{W}^T(\partial_\alpha\mathbf{T}\mathbf{W}^T + \mathbf{W}(\partial_\alpha\mathbf{T})^T) - \tilde{\mathbf{P}}_e^T(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T)^{-1}\tilde{\mathbf{P}}_e(\partial_\alpha\mathbf{T})^T\mathbf{R}(\mathbf{I}-\mathbf{T}\mathbf{W}^T)\Big) \times \left(\mathbf{I}-\mathbf{W}\mathbf{T}^T\right)\mathbf{R}\partial_\beta\mathbf{T} - \mathbf{W}^T(\mathbf{I}-\mathbf{W}\mathbf{T}^T)\mathbf{R}\left(\partial_{\alpha\beta}\mathbf{T} - \partial_\alpha\mathbf{T}\mathbf{W}^T\partial_\beta\mathbf{T}\right)\Big\} \tag{9}$$

reason why computer simulations show, that using spherical coordinates does not reduce the computation time considerably.

4.4. Symmetry of D

Property 4: Let $\mathbf{D}_1, \mathbf{D}_2 = \operatorname{diag}(\pm 1, \ldots, \pm 1)$ be two diagonal matrices with only "1" or "-1" on the main diagonal. If $\mathbf{U}$ is the square matrix of all eigenvectors of $\mathbf{R}$ with $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, then replacing $\mathbf{T}$ with $\mathbf{U}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\mathbf{D}_2$ has no influence on (7).

Proof: First, we would like to point out that a right-multiplication of $\mathbf{T}$ by $\mathbf{D}_2$ is only a special case of property 2. Therefore, we will restrict to the left-multiplication of $\mathbf{T}$ by $\mathbf{U}\mathbf{D}_1\mathbf{U}^T$. One term in (7) is then

$$\begin{aligned}
&\operatorname{tr}\big\{\mathbf{R}\mathbf{U}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{U}\mathbf{D}_1\mathbf{U}^T\mathbf{R}\mathbf{U}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1} \\
&\quad\times \tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{U}\mathbf{D}_1\mathbf{U}^T\mathbf{R}\big\} \\
&= \operatorname{tr}\big\{\mathbf{U}\mathbf{\Lambda}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{U}\mathbf{D}_1\mathbf{\Lambda}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1} \\
&\quad\times \tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{U}\mathbf{D}_1\mathbf{\Lambda}\mathbf{U}^T\big\} \\
&= \operatorname{tr}\big\{\mathbf{U}\mathbf{D}_1\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{D}_1\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e\mathbf{T}^T\big\} \\
&= \operatorname{tr}\big\{\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^T\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e\mathbf{T}^T\big\} \\
&= \operatorname{tr}\big\{\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\left(\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\mathbf{T}\tilde{\mathbf{P}}_e^T\right)^{-1}\tilde{\mathbf{P}}_e\mathbf{T}^T\mathbf{R}\big\}
\end{aligned} \tag{14}$$

In the third line, we used the cyclic shift property $\operatorname{tr}\{\mathbf{A}\mathbf{B}\} = \operatorname{tr}\{\mathbf{B}\mathbf{A}\}$ with $\mathbf{B} = \mathbf{U}\mathbf{D}_1\mathbf{\Lambda}\mathbf{U}^T$. Additionally, $\mathbf{D}_1\mathbf{\Lambda}\mathbf{D}_1 = \mathbf{\Lambda}$ holds as $\mathbf{D}_1\mathbf{D}_1 = \mathbf{I}$ and $\mathbf{D}_1$ and $\mathbf{\Lambda}$ are diagonal matrices. For the last line in (14), we used the identity $\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{R}\mathbf{R}$ and $\operatorname{tr}\{\mathbf{A}\mathbf{B}\} = \operatorname{tr}\{\mathbf{B}\mathbf{A}\}$ with $\mathbf{A} = \mathbf{R}$ once again. ■

Property 4 can be easily interpreted. The right-multiplication with $\mathbf{D}_2$ inverts some columns of $\mathbf{T}$ so that they will point to the opposite direction. As we are transmitting the coordinates along the columns of $\mathbf{T}$, only their direction is important but not their orientation. The interpretation of left-multiplying by $\mathbf{U}\mathbf{D}_1\mathbf{U}^T$ is as follows: First, we do a coordinate transform of $\mathbf{T}$ along the eigenvectors of $\mathbf{R}$ by multiplying with $\mathbf{U}^T$. Afterwards, we change the sign of some rows by $\mathbf{D}_1$. Finally, we reverse the previous coordinate transform by $\mathbf{U}$. This invariance shows that there is a mirror symmetry of (7) along the eigenvectors of $\mathbf{R}$.

5. CONCLUSIONS

A generalized and optimum multiple description coding is considered in this paper. It allows redundant descriptions to be transmitted which offer an improved robustness against erasure channels. Several properties of the distortion measure (7) are proved which help to understand the function of the prefilter. Especially the knowledge of the Hessian matrix (9) allows to efficiently find a solution of the optimization problem.

6. APPENDIX

Let $\mathbf{D} \in \mathbb{R}^{L\times L}$ be an arbitrary diagonal matrix. We will show in this appendix that the identity $\mathbf{D}\tilde{\mathbf{P}}_e^T = \tilde{\mathbf{P}}_e^T\tilde{\mathbf{D}}$ holds, where $\tilde{\mathbf{D}} = \tilde{\mathbf{P}}_e\mathbf{D}\tilde{\mathbf{P}}_e^T \in \mathbb{R}^{\tilde{M}\times\tilde{M}}$. Left-multiplying $\tilde{\mathbf{D}} = \tilde{\mathbf{P}}_e\mathbf{D}\tilde{\mathbf{P}}_e^T$ by $\tilde{\mathbf{P}}_e^T$ yields

$$\tilde{\mathbf{P}}_e^T\tilde{\mathbf{D}} = \tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e\mathbf{D}\tilde{\mathbf{P}}_e^T = \mathbf{D}\tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e\tilde{\mathbf{P}}_e^T = \mathbf{D}\tilde{\mathbf{P}}_e^T \tag{15}$$

as $\tilde{\mathbf{P}}_e^T\tilde{\mathbf{P}}_e$ and $\mathbf{D}$ are diagonal matrices and $\tilde{\mathbf{P}}_e\tilde{\mathbf{P}}_e^T = \mathbf{I}$.

7. REFERENCES

[1] G. Romano, P. Salvo Rossi, and F. Palmieri, "Optimal correlating transform for erasure channels," *IEEE Signal Processing Letters*, vol. 12, no. 10, pp. 677–680, 2005.

[2] SFB 732, "Collaborative research center 732: Incremental specification in context," Available: http://www.uni-stuttgart.de/linguistik/sfb732/.

[3] V. K. Goyal, "Multiple description coding: compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, 2001.

[4] I. Radulovic and P. Frossard, "Multiple description coding with redundant expansions and application to image communications," *EURASIP Journal on Image and Video Processing*, vol. 2007, 2007.

[5] Stefan Uhlich, "Multiple Description Matlab Toolbox," Available: http://www.lss.uni-stuttgart.de/mitarbeiter/uhlich.html.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 1: Estimation Theory*, Prentice-Hall, 1993.

[7] L. E. Blumenson, "A derivation of n-dimensional spherical coordinates," *The American Mathematical Monthly*, vol. 67, no. 1, pp. 63–66, 1960.