NONLINEAR DISTRIBUTED SOURCE–CHANNEL CODING OVER ORTHOGONAL ADDITIVE WHITE GAUSSIAN NOISE CHANNELS

Niklas Wernersson and Mikael Skoglund

School of Electrical Engineering and the ACCESS Linnaeus Center
Royal Institute of Technology (KTH), Stockholm, Sweden

ABSTRACT

The problem of designing simple and energy-efficient nonlinear distributed source–channel codes is considered. By demonstrating similarities between this problem and the problem of bandwidth expansion, a structure for source–channel codes is presented and analyzed. Based on this analysis an understanding about desirable properties for such a system is gained and used to produce an explicit source–channel code which is then analyzed and simulated. It is shown that the code has a substantial gain compared to a linear source–channel code.

***Index Terms*—** Source coding, quantization, channel coding, correlation.

1. INTRODUCTION

Wireless sensor networks are expected to play an important role in tomorrow's sensing systems. One interesting property in these networks is the potential correlation between different sensor measurements which motivates distributed source and channel coding of correlated sources. Most existing literature about how to design practical codes for this problem focus on the case with two sensors and nonlinear source–channel codes, see e.g. [1]. However, in general these approaches are nontrivial to implement for a large number of sensors which has motivated research on linear source-channel codes, see e.g. [2], implementable also for a large number of sensors. The drawback with the linear approach is that it is, in general, suboptimal, see e.g. [3, 4]. Hence, it should be possible to achieve a better performance by allowing nonlinear operations. In this paper, extended in [5], we consider an approach for analog nonlinear distributed source–channel coding, with better performance than linear codes, which can be implemented in the case of a large number of sources.

2. PROBLEM FORMULATION

Consider the problem illustrated in Figure 1. An analog, i.e. continuous-valued, random source sample X with variance σ_x^2 is observed by k separate encoders (sensors) through the noisy observations

$$x_i = x + w_i,\ 1 \leq i \leq k \tag{1}$$

where the W_i's are independent identically distributed (i.i.d.) zero mean Gaussian with variance σ_w^2. Each encoder encodes its own observation x_i by performing an analog mapping, that is $s_i : \mathbb{R} \rightarrow \mathbb{R}$, under the power constraint

$$E[s_i(X_i)^2] \leq P. \tag{2}$$

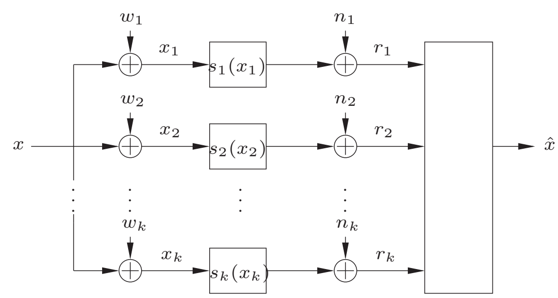


Fig. 1. Structure of the system.

The encoded values,

$$\mathbf{s}(\mathbf{x}) \triangleq (s_1(x_1), s_2(x_2), \ldots, s_k(x_k))^T \tag{3}$$

are transmitted over k orthogonal AWGN channels, created by using e.g. TDMA, FDMA or CDMA, and the decoder estimates x based on the received values

$$\mathbf{r} = \mathbf{s}(\mathbf{x}) + \mathbf{n} \tag{4}$$

where $\mathbf{N}$ is i.i.d. memoryless Gaussian distributed with covariance matrix $\sigma_n^2 I$. Hence, the decoding is performed as

$$\hat{x} = \hat{x}(\mathbf{r}) \tag{5}$$

and the objective is to minimize the expected mean square error (MSE) $E[(X - \hat{X})^2]$. The main focus of this paper is on how to design $\mathbf{s}(\mathbf{x})$.

3. DISCUSSION AND PROPOSED SCHEME

We will in Subsections 3.A–B discuss two important special cases of the problem described in Figure 1. Understanding for these special cases leads to insight about how to design the encoding function $\mathbf{s}(\mathbf{x})$ for the general case. Based on this insight we present and analyze a structure for $\mathbf{s}(\mathbf{x})$ in Subsections 3.C–D. Finally, based on the derived results we propose an explicit scheme for $\mathbf{s}(\mathbf{x})$ in Subsection 3.E.

3.1. $\sigma_w^2 > 0$ and $\sigma_n^2 \rightarrow 0$

For the case when $\sigma_w^2 > 0$ and $\sigma_n^2 \rightarrow 0$ the AWGN channels are approaching ideal, i.e. noiseless, channels and r_i will approach $s_i(x_i)$. This means that given the *linear encoding* strategy,

$$s_i(x_i) = \sqrt{\frac{P}{\sigma_x^2 + \sigma_w^2}} x_i,\ 1 \leq i \leq k, \tag{6}$$

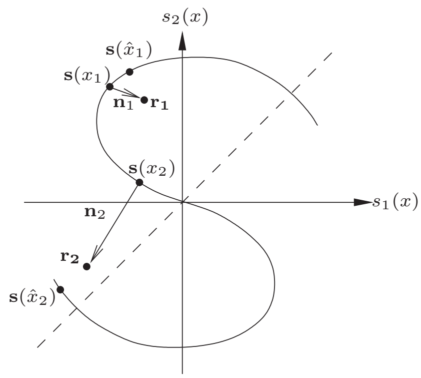


Fig. 2. x_1 illustrates a 'small' decoding error and x_2 illustrates a 'large' decoding error.

the decoder will get access to the noisy observations $\{x_i\}_{i=1}^k$ (since the channel is close to perfect). Given these observations we could theoretically perform the best possible estimation based on the noisy observations as given by the Cramer–Rao lower bound, see e.g. [6]. It is clear that there will be no way to obtain a better performance, since that would require better sensor observations, and we can therefore conclude that the linear coding strategy described above is approaching the optimal strategy when $\sigma_n^2 \to 0$.

3.2. $\sigma_w^2 = 0$ and $\sigma_n^2 > 0$

In the case when $\sigma_w^2 = 0$ and $\sigma_n^2 > 0$ we will get $x_i = x \ \forall i$ which we will write as $\mathbf{s}(\mathbf{x}) = \mathbf{s}(x)$. This problem is equivalent to the problem often referred to as the *bandwidth expansion problem*, see e.g. [7] and the references therein. (See also [8] for the connection between distributed source–channel coding and *bandwidth compression*.) It is well known for the bandwidth expansion problem that when the source is i.i.d. zero-mean Gaussian and $k = 1$, linear encoding is optimal under the assumption that the decoder knows the source and noise variances. However, when $k > 1$ this is no longer true and nonlinear encoding can have superior performance compared to linear encoding strategies, see e.g. [9]. One of the reasons for this is that a linear encoding function $\mathbf{s}(x)$ only uses a one dimensional subspace of the available channel space. More efficient mappings would use a higher number of the available channel space dimensions. An example of this is illustrated in Figure 2 where $k = 2$ is assumed. By using nonlinear encoding functions, illustrated by the solid 'S–shaped' curve $\mathbf{s}(x)$, we are able to better fill the channel space than when using linear encoding functions, represented by the dashed curve. As long as we decode to the right fold of the curve, illustrated by sample x_1 in the figure, a longer curve essentially means a higher resolution when estimating x, i.e. a better estimate. However, decreasing the SNR will at some point result in that different folds of the curve will lie too close to each other and the decoder will start making large decoding errors, illustrated by sample x_2 in the figure. Decreasing the SNR below this threshold will therefore significantly deteriorate the performance.

3.3. Objective

Based on the intuition from these two special cases we can conclude that for the problem considered in this paper, where both $\sigma_w^2 > 0$ and $\sigma_n^2 > 0$, good encoding functions $\mathbf{s}(\mathbf{x})$ should take both these aspects into consideration. Our objective in this paper is to analyze and design nonlinear encoding functions $\mathbf{s}(\mathbf{x})$ of the type illustrated in Figure 2.

3.4. Analysis

In order to gain understanding for the use of nonlinear encoding functions $\mathbf{s}(\mathbf{x})$ we make a performance analysis under the assumptions that σ_w^2 and σ_n^2 are small. We also assume, in this subsection, that all encoding functions $s_i(x_i)$ are continuous and differentiable and that the curve $\mathbf{s}(x)$ is appropriately designed such that no large decoding errors occur under the assumed noise variances. Let us start by studying a certain encoded observation x and the resulting estimate $\hat{x} = x + z$, with z representing the estimation error. Under the above assumptions, i.e. small noise variances, also z will be small and hence

$$\mathbf{s}(\hat{x}) = \mathbf{s}(x + z) \approx \mathbf{s}(x) + z\mathbf{s}'(x). \tag{7}$$

where

$$\mathbf{s}'(x) \triangleq \left(\frac{\mathrm{d}}{\mathrm{d}x}s_1(x), \frac{\mathrm{d}}{\mathrm{d}x}s_2(x), \dots, \frac{\mathrm{d}}{\mathrm{d}x}s_k(x)\right)^T. \tag{8}$$

Now consider the decoder. It is well known that in order to minimize the MSE the decoder should be implemented as

$$\hat{x}(\mathbf{r}) = E[X|\mathbf{r}]. \tag{9}$$

This function will however, in general, be difficult to implement. We will therefore consider the suboptimal maximum likelihood (ML) decoder. Since σ_w^2 is small we get

$$\mathbf{s}(\mathbf{x}) \approx \mathbf{s}(x) + \mathrm{diag}(\mathbf{s}'(x))\mathbf{w} \tag{10}$$

which gives

$$\mathbf{r} = \mathbf{s}(\mathbf{x}) + \mathbf{n} \approx \mathbf{s}(x) + \mathrm{diag}(\mathbf{s}'(x))\mathbf{w} + \mathbf{n}. \tag{11}$$

We approximate the ML decoder as

$$\hat{x}(\mathbf{r}) = \arg\max_x p(\mathbf{r}|x) \approx \arg\min_x \|\mathbf{s}(x) - \mathbf{r}\|^2 \tag{12}$$

where $p(\cdot|\cdot)$ denotes the transition pdf from x to $\mathbf{r}$. Hence, the decoding function corresponds to decoding $\mathbf{r}$ to the closest point on the curve $\mathbf{s}(\hat{x})$. However, for small values of $|x - \hat{x}|$ the curve $\mathbf{s}(\hat{x})$ is approximately linear and parallel to $\mathbf{s}'(x)$. This means that the decoder will remove the noise contributions orthogonal to $\mathbf{s}'(x)$ and we get

$$\mathbf{s}(\hat{x}) \approx \mathbf{s}(x) + \frac{(\mathrm{diag}(\mathbf{s}'(x))\mathbf{w} + \mathbf{n}) \cdot \mathbf{s}'(x)}{\|\mathbf{s}'(x)\|} \frac{\mathbf{s}'(x)}{\|\mathbf{s}'(x)\|} \tag{13}$$

where the dot product describes the projection of the added noise onto the vector $\mathbf{s}'(x)$. Expanding the dot product we get

$$\begin{aligned}(\mathrm{diag}(\mathbf{s}'(x))\mathbf{w} + \mathbf{n}) \cdot \mathbf{s}'(x) \\ = (\mathrm{diag}(\mathbf{s}'(x))\mathbf{w}) \cdot \mathbf{s}'(x) + \mathbf{n} \cdot \mathbf{s}'(x) \triangleq w + n\end{aligned} \tag{14}$$

where

$$W \sim \mathcal{N}(0, \sigma_w^2 \sum_{i=1}^k s_i'(x)^4)$$

$$N \sim \mathcal{N}(0, \sigma_n^2 \sum_{i=1}^k s_i'(x)^2).$$

From (7) and (13) we identify

$$z \approx \frac{w+n}{\|\mathbf{s}'(x)\|^2} \tag{15}$$

and hence

$$\begin{aligned} E[(x-\hat{X})^2] = E[Z^2|x] &\approx \frac{E[W^2|x]}{\|\mathbf{s}'(x)\|^4} + \frac{E[N^2|x]}{\|\mathbf{s}'(x)\|^4} \\ &= \sigma_w^2 \frac{\sum_{i=1}^k s_i'(x)^4}{\|\mathbf{s}'(x)\|^4} + \sigma_n^2 \frac{1}{\|\mathbf{s}'(x)\|^2}. \end{aligned} \tag{16}$$

From this we conclude

$$E[(X-\hat{X})^2] \approx \int f(x) \left[\sigma_w^2 \frac{\sum_{i=1}^k s_i'(x)^4}{\|\mathbf{s}'(x)\|^4} + \sigma_n^2 \frac{1}{\|\mathbf{s}'(x)\|^2} \right] \mathrm{d}x \tag{17}$$

where $f(x)$ is the pdf of x. The second term in (17) is the MSE contribution from the channel noise. (This term was also derived in [9] for the bandwidth expansion case.) It tells us that we should aim for stretching the curve as much as possible, like stretching a rubber band, keeping in mind the constraint (2) at the same time as we also keep a high enough distance between different folds of the curve preventing large decoding errors. In order to stretch the curve it needs to turn in different directions which occurs when $s_i'(x) \neq s_j'(x)$ for some $i \neq j$. If we instead study the first part of (17), which is the MSE contribution from the observation noise, it is straightforward to show that it is minimized when $s_1'(x) = s_2'(x) = \cdots = s_k'(x)$. This indicates a linear system. Hence, from (17) we understand the tradeoff between optimizing the system for being robust to the channel noise and for being robust to the observation noise: If we want to combat the channel noise we should create nonlinear curves $\mathbf{s}(\mathbf{x})$. On the other hand, if we want to combat the observation noise linear encoding functions will be more appropriate. This further tells us that in a situation where σ_w^2 is low, i.e. there is high correlation between the observations, there is a lot to be gained by designing nonlinear encoding functions $\mathbf{s}(\mathbf{x})$. For the opposite case, i.e. low correlation, we can expect less gain from using nonlinear encoding functions $\mathbf{s}(\mathbf{x})$.

3.5. Proposed Scheme

Based on the analysis in Section 3.4 we concluded that linear functions $\mathbf{s}(\mathbf{x})$ are good with respect to the observation noise but inefficient with respect to the channel noise. Therefore, in order to produce a flexible code able to handle both observation and channel noise, we propose a piecewise linear encoding function $\mathbf{s}(\mathbf{x})$ as follows

$$s_i(x_i) = \begin{cases} \alpha x_i & 1 \le i \le k_0 \\ \alpha(x_i - \Delta \left\lfloor \frac{x_i}{\Delta} \right\rceil) & k_0 < i \le k \end{cases} \tag{18}$$

where $\lfloor \cdot \rceil$ denotes rounding to the nearest integer, α will control the power usage and Δ will for $i > k_0$ control the length of each period in the periodic sawtooth function $s_i(x_i)$. Hence, we allow noncontinuous and nondifferentiable functions which was not the case in the analysis. The reason is that this results in a system where $s_1'(x) = s_2'(x) = \cdots = s_k'(x)$, except at the discontinuities, which is desirable with respect to the observation noise. At the same time we get a nonlinear system, better able to use the available channel space, which is desirable with respect to the channel noise. The drawback is that the approximation in (11) is violated making the decoder (12) inefficient. We therefore modify the decoding function as follows:

1. Create the ML estimate of x based on the linear encoding functions:
$$\hat{x}_{k_0} = \frac{1}{k_0} \sum_{i=1}^{k_0} \frac{r_i}{\alpha}. \tag{19}$$
2. Assume that $|\hat{x}_{k_0} - x_i - n_i/\alpha| \le \Delta/2$ for $k_0 < i \le k$ and create the ML estimates
$$\begin{aligned} \hat{x}_i(r_i) = \arg\min_{x_i} &\left((s_i(x_i) - r_i)^2 \right| \\ &x_i \in \{|\hat{x}_{k_0} - x_i| \le \Delta/2\} \big). \end{aligned} \tag{20}$$
This function tries to predict the removed part $\Delta \left\lfloor \frac{x_i}{\Delta} \right\rceil$ from (18) based on the derived $\hat{x}_{k_0}$.
3. Based on this, create the final estimate of x as
$$\hat{x} = \frac{1}{k} \left(\sum_{i=1}^{k_0} \frac{r_i}{\alpha} + \sum_{i=k_0+1}^{k} \hat{x}_i(r_i) \right). \tag{21}$$

Let us now analyze the power consumption. Note that the power used by the nonlinear encoding functions will be less than the power used by the linear encoding functions. We define the normalized average power consumption as

$$P(\Delta, k_0) = \frac{1}{k\alpha^2} \left(k_0 E[s_I(X_I)^2] + (k-k_0) E[s_J(X_J)^2] \right) \tag{22}$$

where we assume $I \le k_0 < J \le k$. (The reason for dividing with α^2 is that we want $P(\Delta, k_0)$ to represent the change in power consumption due to Δ and k_0 and not due to the scaling factor α.) By performing timesharing the sensors could use the linear encoding function for a fraction k_0/k of the available time slots and then use the nonlinear encoding functions the rest of the time. Hence, $P(\Delta, k_0)$ can be seen as the average power used by each sensor when $\alpha = 1$. We therefore choose

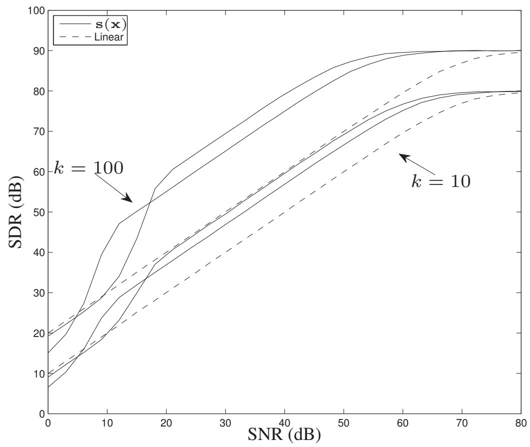
$$\alpha = \sqrt{\frac{P}{P(\Delta, k_0)}} \tag{23}$$

in order to fulfill (2). In [5] we derive MSE for the proposed scheme (18) as a function of Δ and k_0. This gives us the possibility to optimize the choice of Δ and k_0 for a certain set of noise parameters. Lack of space prevents us from including these derivations in this paper.

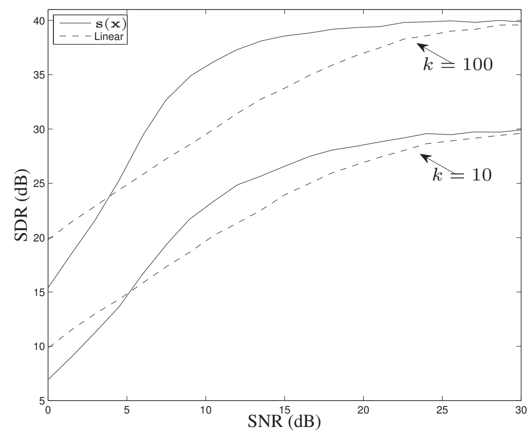
4. SIMULATIONS

In Figure 3 we present simulation results for the proposed code (18). In both simulations we used a zero-mean i.i.d. Gaussian distribution for X with variance σ_x^2. We measure the performance in SDR $\triangleq (\sigma_x^2 + \sigma_w^2)/E[(X-\hat{X})^2]$ versus SNR $\triangleq P/\sigma_n^2$ and the correlation is measured as $\rho \triangleq \sigma_x^2/\sigma_w^2$. In order to design the systems for different noise levels we have used the formula from [5] to optimize the choices of Δ and k_0. In the simulations we let ρ equal what the system has been designed for, but we vary the SNR in order to study the effects of SNR mismatch. We present results for two cases: $k = 10$ and $k = 100$. The performance of the linear system is shown by the dashed lines and the performance of the proposed scheme $\mathbf{s}(\mathbf{x})$ is shown by the solid lines.

In Figure 3(a) the systems have been optimized for an environment corresponding to the case described in Section 3.2, i.e. close to



(a)



(b)

Fig. 3. Systems for $k = 10$ and $k = 100$ designed for (a) $\rho = 70$dB and SNR$=$ 10dB as well as SNR$=$ 20dB (b) $\rho = 20$ dB and SNR$=$ 10dB. In all simulations the 'true' ρ has been used but the SNR is varied.

bandwidth expansion which occurs when ρ has a high value. We have here included two designs for $\rho = 70$ dB, namely SNR$=$ 10 dB and SNR$=$ 20 dB. Hence, one of the systems is optimized for a better channel meaning that different folds of $\mathbf{s}(x)$ should be packed closer to each other. This will give a better performance for high SNRs but the code will also break down faster when the SNR is decreased. This is clearly visible in the figure. The behavior of the designed systems follows the predicted behavior from (17): For low SNRs the nonlinear codes break down and the linear system is the better one. However, increasing the SNR above a certain threshold makes the probability of large decoding errors small and the code will start work as a bandwidth expansion code. Decreasing σ_n^2 above this threshold with some factor, hence increasing the SNR, will lead to the same factor of decrease in the second term of (17). This behavior is clearly visible in the figure. Increasing the SNR even more will at some point make the first term of (17) the dominant one and the performance reaches a saturation level.

Finally, we consider the, for this paper, more interesting case where we optimize systems with fair amounts of both observation and channel noise. In Figure 3(b) we have designed systems for $\rho =$ 20 dB and SNR$=$ 10 dB. Also here we simulate for the designed ρ but we vary the SNR. It is clear that there will be a large gain over the linear system except at low SNRs, where the source–channel code breaks down, and at high SNRs, where there is no gain in using nonlinear codes.

5. CONCLUSIONS

We have explained the similarities between the problem of distributed source–channel coding and the problem of bandwidth expansion. Based on this we have presented and analyzed a suitable structure for distributed source–channel codes. This analysis gives us insight into desirable properties for such a system. Based on this understanding an explicit code is presented, analyzed and simulated. A large gain was obtained from the nonlinear code compared to the linear code.

References

[1] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, March 2003.

[2] S. Cui, J. Xiao, A. Goldsmith, Z. Luo, and V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, September 2007.

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

[4] J. J. Xiao and Z. Q. Luo, "Multiterminal source-channel communication over an orthogonal multiple-access channel," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3255–3264, September 2007.

[5] N. Wernersson and M. Skoglund, "Nonlinear coding and estimation for correlated data in wireless sensor networks," *IEEE Transactions on Communications*, 2008, submitted.

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

[7] V. Vaishampayan and S. I. R. Costa, "Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1658–1672, July 2003.

[8] N. Wernersson and M. Skoglund, "Nonlinear distributed source–channel coding over the gaussian multiple access channel," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, submitted.

[9] D. J. Sakrison, *Transmission of Waveforms and Digital Information*. Wiley, 1968.