LOW-COMPLEXITY CODING FOR THE CEO PROBLEM WITH MANY ENCODERS

Gerhard Maierbacher and João Barros

Department of Computer Science
University of Porto, Portugal
URL: `http://www.dcc.fc.up.pt/~barros`

Motivated by a sensor networking application, we consider the CEO problem, in which noisy observations of a common source process are separately encoded and sent to a joint receiver. The goal of the decoder is to produce the best possible reconstruction of the source subject to some distortion criterion. Focusing on the quadratic Gaussian case, for which the rate-distortion function has recently been determined, we present practical coding schemes that are feasible for a very large number of distributed encoders, allowing a desirable trade-off between rate-distortion performance and algorithmic complexity.

1 INTRODUCTION

1.1 The quadratic Gaussian CEO Problem

Consider the following multiterminal source coding problem proposed in [7]. A CEO (**C**hief **E**xecutive **O**fficer) is interested in some data U_0, which is not directely accessible. Therefore the CEO employs a team of M agents, who observe independently corrupted versions of U_0. Because the estimation of U_0 is only one among many pressing matters to which the CEO must attend, the combined rate at which the agents may communicate information about their observation to the CEO is limited, say to R_Σ bits per second. The agents are not allowed to convene. The CEO's task is to estimate U_0 with the smallest possible average distortion D.

The described setup offers a reasonable model for sensor networks, in which physical data is collected and encoded independently by a large number of geographically separated sensor nodes, possibly observing different characteristics of the same source (e.g., multispectral imaging). The encoded data must then be sent over a channel of limited capacity to a remote station for data fusion and further processing.

We speak of the quadratic Gaussian CEO problem (see Fig. 1), when the distortion measure is the mean square error, and both the targeted data U_0 and the additive observation noises N_m, $i = 1, \ldots, M$. are Gaussian distributed [2].

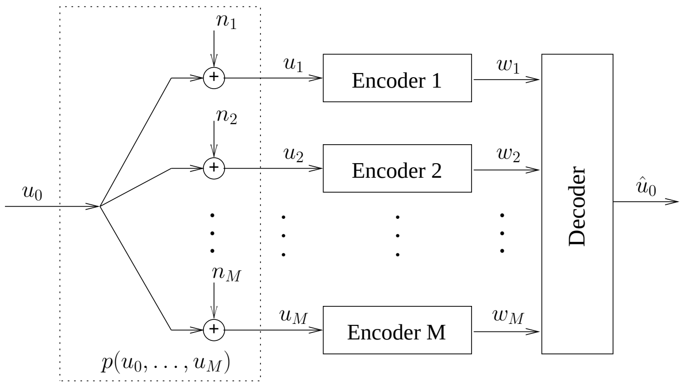

Figure 1: Problem setup for the quadratic Gaussian CEO problem. M separate encoders observe different versions $u_i = u_0 + n_i$ of the Gaussian source u_0 corrupted by Gaussian noises n_i, $i = 1 \ldots M$. The decoder must produce the reconstruction $\hat{u}_0$ based on the received messages $w_1 \ldots w_M$.

The (sum) rate-distortion function for this relevant special case, was studied in [3], which presents an upper bound found to be tight for noise processes with identical variance. However, the proof thereof is non-constructive and design of practical codes for this problem is still at an infant stage. Seeking to develop low-complexity coding methods for the quadratic Gaussian CEO problem with a very large number(i.e. $M > 100$) of parallel observations, we make the following contributions:

- a feasible decoder that relies on a factor graph description [4, 1] of the source statistics, the noisy observations and the received messages; both its computational and its space complexity grow linearly with the number of encoders;
- a low-complexity encoding stage based on scalar quantization and optimized index assignments that exploit the statistical dependencies of the data to provide a wide range of achievable rate-distortion points and very low coding delay;

- two different methods to obtain appropriate index-assignments: (i) generalized index-reuse [8] and (ii) random puncturing.
- analytic and numerical performance results that underline the effectiveness and efficiency of our low-complexity coding strategies.

This rest of the paper is organized as follows: In Section 2 we introduce our notation and provide a formal statement of the problem. Then, Section 3 explains our decoding procedure based on the sum-product algorithm, as well as the design of the index-assignment stage for the distributed scalar quantizers. Finally, Section 4 presents some numerical results.

2 PROBLEM SETUP

2.1 Notation

We begin by introducing some notation: Random variables are denoted by capitalized letters, e.g. A, with the corresponding lowercase symbol a, representing a possible realization. Vectors are denoted with small bold symbols $\mathbf{a}$, whereas capital bold letters $\mathbf{A}$ are used for matrices. Calligraphic letters, e.g. $\mathcal{A}$, denote sets, representing for instance the alphabet of a random variable or a set of functions. The covariance is defined by $Cov\{\mathbf{a}, \mathbf{b}\} = E\{\mathbf{a}\mathbf{b}^T\}E\{\mathbf{a}\}E\{\mathbf{b}\}^T$, where $E\{\cdot\}$ is the expectation operator. An N-dimensional random variable with realizations $\mathbf{a} \in \mathbb{R}^N$ is Gaussian distributed with mean $\boldsymbol{\mu} = E\{\mathbf{a}\}$ and covariance matrix $\boldsymbol{\Sigma} = Cov\{\mathbf{a}, \mathbf{a}\}$, when its probability density function (PDF) $p(\mathbf{a})$ is given by

$$p(\mathbf{a}) = \frac{1}{(2\pi|\boldsymbol{\Sigma}|)^{1/2}} \exp(-\frac{1}{2}(\mathbf{a}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{a}-\boldsymbol{\mu})). \tag{1}$$

Such a PDF is denoted as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

2.2 System Description

Our basic system model is illustrated in Figure 1. Let $u_0(t)$ be the output of a continous-valued Gaussian source U_0 at time instant t. For $m = 1, 2, ..., M$ let $u_m(t)$ denote noisy observations of $u_0(t)$, which are corrupted by additive noise, i.e. $u_m(t) = u_0(t) + n_m(t)$. The noise samples $n_m(t)$ are generated by Gaussian noise processes N_m, statistically independent over m. For simplicity, we drop the time variable t and consider only one time step. Thus $u_m = u_0 + n_m$, where the probability density functions (PDF's) of u_0 and n_m can be described by one-dimensional Gaussian distributions, $\forall m \in \{1, 2, ..., M\}$. We assume, that u_0 is distributed by $\mathcal{N}(s_0, \sigma_0^2)$ with mean s_0 and variance σ_0^2, n_m is distributed by $\mathcal{N}(l_m, \lambda_m^2)$ with mean l_m and variance λ_m^2.

The encoders are "cheap" devices consisting of a scalar quantizer and a mapping stage. Each encoder m converts its scalar observation u_m to a quantization-index $i_m \in \mathcal{I}_m = \{1, 2, ..., L_m\}$, corresponding to L_m reconstruction values $\tilde{u}_m(i_m) \in \mathbb{R}$. The quantization indices of all encoders are represented by the vector $\mathbf{i} = [i_1, \cdots, i_M]^T$. The mapping-stage is used to assign the quantizer-outputs i_m to codewords indexed by $w_m \in \mathcal{W}_m = \{1, 2, ..., K_m\}$. We refer to this mapping procedure as an *index assignment*. The codeword-indices of all encoders are represented by the vector $\mathbf{w} = [w_1, \cdots, w_M]^T$. In the later part of this paper, we will have need for the binary representation of the codewords at encoder m, denoted as $\mathbf{b}_m = [b_{m,1}...b_{m,Q_m}]$, $b_{m,i} \in \{0, 1\}$ with $Q_m = \lceil \log_2(K_m) \rceil$.

The decoder uses the received codewords of all encoders $\mathbf{w}$ to form an estimate $\hat{u}_0$ of the current source sample u_0. Its performance is evaluated by a quadratic distortion measure.

Since all encoders gather data about the same source process, the encoder observations have to be necessarily correlated. This correlation can be used to improve the decoding result at a given transmission rate for each encoder m. Generally speaking, we can exploit the correlations prior transmission by eliminating redundancy within the encoder data (Slepian-Wolf coding, [6]) or at the decoder side by using the joint statistics of the source and all observations to improve the estimates fidelity. Both strategies are applied within our system, to optimize the overall rate/distortion performance. Notice, that the index assignment stage is capable of performing a simple form of Slepian-Wolf coding, by chosing appropriate index assignments, where $K_m < L_m$.

3 CODE CONSTRUCTION

3.1 Decoding Stage

The key aspect that renders the CEO problem more tractable than other multiterminal source coding problems is the fact that all noisy sensor observations $\mathbf{U} = (U_1, ..., U_M)$ are statistically independent given the source-process U_0. The special structure of the problem leads to a Markov chain property of the form $W_k - I_k - U_k - U_0 - U_l - I_l - W_l$, $\forall\, k, l \in \{1, ..., M\}$, where $k \neq l$. This induces a natural factorization of the joint probability function according to

$$p(u_0, \mathbf{w}) = p(u_0) \cdot p(\mathbf{w}|u_0) = p(u_0) \cdot \prod_{m=1}^{M} p(w_m|u_0), \tag{2}$$

which can be represented by a factor-graph [4].The main advantage of this graphical representation is that we can use the sum-product algorithm [4] to obtain decoding estimates that minimize the mean squared error (MMSE) $E\{(u_0 - \hat{u}_0)^2\}$. It is well known that the optimal decoder in this case is the conditional mean estimator[5] given by

$$\hat{u}_0^{opt}(\mathbf{w}) = E\{U_0|\mathbf{w}\} = \int_{u_0=-\infty}^{+\infty} u_0 \cdot p(u_0|\mathbf{w}) \; du_0. \tag{3}$$

To ease the decoding computations, we approximate this integral by the sum

$$\hat{u}_0(\mathbf{w}) = \sum_{i_0=1}^{L_0} \tilde{u}_0(i_0) \cdot p(i_0|\mathbf{w}), \tag{4}$$

where we use a fixed number of reconstruction values of the source u_0 denoted by $\tilde{u}_0(i_0)$ and indexed by $i_0 \in \mathcal{I}_0 = \{1, ..., L_0\}$. Exploiting further the one-to-one mapping of the indices i_0 to the reconstruction values $\tilde{u}_0$, the factorization in (2) can be modified to yield the decoding rule

$$\hat{u}_0(\mathbf{w}) = c(\mathbf{w}) \cdot \sum_{i_0=0}^{L_0-1} \tilde{u}_0(i_0) \cdot p(i_0) \prod_{m=1}^{M} p(w_m|i_0), \tag{5}$$

where $c(\mathbf{w}) = 1/p(\mathbf{w})$ is the normalization constant. The probabilities $p(w_m|i_0)$ can be determined numerically in a one-time procedure prior to decoding.

It is possible to show that the algorithmic complexity of the proposed decoding scheme scales according to $\mathcal{O}(ML_0)$. The space complexity (in terms of memory requirement for the a-priori information) grows $\mathcal{O}(MKL_0)$. In other words both types of decoding complexity grow linearly in the number of sensors.

3.2 Index Assignment Strategies

The index assignment stage, which is represented by a mapping function f_m such as $w_m = f_m(i_m), \quad \forall m \in \{1, \cdots, M\}$, is able to eliminate the available redundancy by taking $K_m < L_m$ codewords (i.e. less than the number of quantization levels thus reducing the coding rate). In the following, we propose two conceptually different methods - index reuse and random puncturing.

The proposed index reuse method is inspired by [8], originally devised for two correlated sources. The basic idea is to reduce the number of possible output codewords in an iterative fashion, computing in every step the resulting end-to-end distortion. Starting with a one-to-one mapping between the indices i_m and the codewords w_m, i.e. $f_m = i_m$, $\forall m \in \{1, ..., M\}$, the algorithm subsequently modifies the mapping functions f_m of all encoders m, by merging two codewords (or, equivalently, the quantization cells) to a single new codeword. This is done repeatedly until the targeted number of codewords K_m is reached. In each step, the algorithm choses the merging possibility that yields the minimum possible average distortion. We note that the search algorithm is not optimal due to the single-step nature of the optimization.

A detailed formulation of the whole procedure can be found in Algorithm 1, which assumes $L_m = L$ and $K_m = K, \ \forall \ m \in \{1, ..., M\}$. We define a merging function $f^*_m = g(f_m, a, b)$, where f^*_m denotes the resulting mapping function. To obtain f^*_m from f_m, where $a < b$, the following assignment has to be performed for $l = 1, ..., L_m$:

$$f^*_m(l) = \begin{cases} a & \text{, for } f_m(l) = a \text{ or } f_m(l) = b \\ f_m(l) - 1 & \text{, for } f_m(l) > b \\ f_m(l) & \text{, otherwise.} \end{cases} \quad (6)$$

In each step the distortion can be approximated according to

$$D = \sum_{\forall \mathbf{w}} \sum_{i_0=1}^{L_0} p(i_0, \mathbf{w}) \cdot (\tilde{u}_0(i_0) - \hat{u}_0(\mathbf{w}))^2, \quad (7)$$

where $\hat{u}_0(\mathbf{w})$ is obtained by (5). Due to the complexity of this computation, the algorithm is only feasible for a small number of encoders. A straightforward way to decrease the computational complexity is to form clusters of encoders and optimize each cluster separately, as exemplified in the next section.

Alternatively, we propose a very simple method to reduce the coding rate: ommit (i.e. puncture) some of the bits of the binary codewords $\mathbf{b}_m$ according to a fixed, randomly chosen puncturing pattern known both to the encoders and to the decoder. The result is an equivalent index assignment, which can be easily incorporated in the decoding algorithm described in Section 3.1.

Algorithm 1 Optimization Algorithm for m encoders

Initialization

- start with one-to-one mapping
  $f_m \leftarrow [1, \cdots, L]$, $\forall\, m \in \{1, ..., M\}$
- set initial nr. of codewords
  $k \leftarrow L$

Main Loop

```
repeat
  for m = 1 to M do
    • set distortion to maximum
        D ← 1
    for a = 1 to (k − 1) and b = a to k do
      • merge cell a and cell b within f_m
          f*_m = g(f_m, a, b)
      • calculate resulting overall distortion
          D* = D(F_\m, f*_m)
      if (D* < D) then
        • remember mapping and distortion
            D ← D*
            f_m ← f*_m
      end if
    end for
  end for
  • cells are merged and new mappings F found ⇒
  nr. of codewords is reduced by one
      k ← k − 1
until k = K; i.e. targeted nr. of codewords reached
```

4 RESULTS

The end-to-end distortion of the proposed schemes can be determined analytically up to a numerical integration of the Gaussian distribution. We now present numerical results for a scenario of $M = 128$ encoders, where identical quantizers and identical transmission rates are considered. The source process is Gaussian distributed with variance $\sigma_0^2 = 1$ and mean $s_0 = 0$, the noise processes are Gaussian distributed with variance $\lambda_m^2 = \lambda^2$ and mean $l_m = 0$, $\forall m \in \{1, ..., M\}$. All scalar quantizers are Lloyd-Max optimized to minimize the mean squared error within the observations. The decoder uses a discrete source model of precision $L_0 = 64$ based on uniform quantization, which is optimized for the source statistics. The sum-product algorithm is used for decoding, where the probabilities $p(i_0)$ and $p(i_m|i_0)$ were determined using Monte Carlo simulation. In the case of index reuse, the optimization algorithm is run over a subset of four encoders. Thus, we obtain four index assignments, which were repeatedly assigned to all encoders (symmetric scenario). Notice, that the index assignment yielding the best possible performance was chosen for the experiments (e.g. $K = 2$ codewords may be obtained from quantizers of $L = 4$ or $L = 16$). The bit puncturing approach builds on standard Gray mappings. Again, we took the s cheme yielding the best possible performance. Opting for $N = 10000$ samples we computed the

Quantizer Q [bit]	Low-Noise: σ_0=1,λ=0.1 1	2	3	4	High-Noise: σ_0=1,λ=0.5 1	2	3	4
$\mathrm{SNR}_{\mathrm{R/D}}$[dB]	29.75	30.78	31.00	31.06	22.79	23.81	24.03	24.08
$\mathrm{SNR}_{\mathrm{Dec}}$[dB]	10.14	19.80	26.63	28.74	15.61	22.10	23.45	23.59
$\mathrm{SNR}_{\mathrm{IR}}$[dB]	22.89	27.58	28.61	28.74	19.59	22.32	23.37	23.59
$\mathrm{SNR}_{\mathrm{BP}}$[dB]	24.51	27.26	28.28	28.75	17.58	20.87	22.50	23.57

Table 1: Numerical results for $M = 128$ encoders for low-noise (left) and high-noise (right) scenario.

output signal-to-noise ratio $\mathrm{SNR} = 10 \log_{10} \frac{\sigma_0^2}{D}$ to quantify the overall fidelity. The results are summarized in Table 1, which includes the theoretically optimal value $\mathrm{SNR}_{\mathrm{R/D}}$ as given by the sum-rate distortion function in [3]. $\mathrm{SNR}_{\mathrm{Dec}}$ is the output SNR of the system without index assignments, i.e. determined by the decoder alone, where $\mathrm{SNR}_{\mathrm{IR}}$ and $\mathrm{SNR}_{\mathrm{BP}}$ is the output SNR when index reuse (IR) or bit puncturing (BP) is applied. Our experiments reveal that when the observation noise level exceeds a certain threshold the index-reuse method can outperform the random puncturing scheme. As expected, for high rates both methods approach the theoretical rate-distortion limit.

REFERENCES

[1] J. Barros and M. Tüchler. Scalable decoding on factor graphs: a practical solution for sensor networks. To appear in the IEEE Transactions on Communications , 2005.

[2] Harish Viswanathan, Toby Berger. The quadratic Gaussian CEO problem. *IEEE Trans. Inform. Theory*, 43:1549–1559, 1997.

[3] Jun Chen, Xin Zhang, Toby Berger, Stephen B. Wicker. An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem. *Special Issue of JSAC, On Fundamental Performance of Wireless Sensor Networks*, 2004.

[4] F. R. Kschischang, B. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. on Inform. Theory*, 47(2):498–519, 2001.

[5] H. V. Poor. *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1994.

[6] D. Slepian and J. K. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Trans. Inform. Theory*, IT-19(4):471–480, 1973.

[7] T. Berger, Z. Zhang and H. Viswanathan. The CEO problem. *IEEE Trans. Inform. Theory*, 42:887–902, 1996.

[8] Thomas J. Flynn, Robert M. Gray. Encoding of correlated observations. *IEEE Trans. Inform. Theory*, IT-33:773–787, 1987.