Optimising the combined source and channel coding of a discrete communication system

A. Miri, E. Hons and A.K. Khandani

Abstract: The problem of optimising the structure of the encoder/decoder pair in a discrete communication system, with an additive distortion measure, is formulated in terms of a quadratic programming (QP) problem. This new formulation benefits from the following special features: it optimises the joint effects of the source/channel coding on the end-to-end distortion; and the encoder and the decoder structures are not restricted to being the inverse of each other. A method which obtains an ε-minimiser approximation of an optimum point of a general QP problem is discussed. Some simulation results based on this method are also given.

1 Introduction

Vector quantisation (VQ) has been a popular and effective technique which addresses the problem of transmitting a source subject to a fidelity criterion. An important part of a VQ's design is its robustness in the presence of channel noise. In this paper, we are concerned with the low-complexity design of an overall system, which involves no explicit error control to deal with channel errors, and does not suffer delays due to channel decoding. It has been shown [1–4] that in such systems an appropriate selection of the mapping between the source and the channel symbols can reduce the effect of channel errors, which can otherwise be severe. This mapping is conventionally called the 'index-assignment problem'. Two approaches are typically taken to this problem: one is to design a VQ source coder for a noiseless channel, and add to it an optimised index assignment. The other approach incorporates the effects of channel error directly into the design of the VQ. These VQs are in a sense optimised for a given channel and often are referred to as 'channel optimised VQ'. Although channel optimized VQs typically outperform those designed using index assignment techniques, they require longer training time and the exact knowledge of channel characteristics, which may not always be available. In this paper we adopt the index-assignment approach. Several index-assignment methods have been studied in the literature. De Marca and Jayant [1] introduced an iterative search algorithm for designing index assignments for scalar quantisers, which was extended to vector quantisation in [5]. For binary-symmetric channels and certain special sources and quantisers, analytical results have been obtained [6–9]. Using a 'greedy' index assignment method, some numerical bounds for noisy channel vector quantisation were presented in [10]. Zegger and Manzella [11] presented a random coding argument for the selection of the index assignment resulting in certain bounds on the system performance. Knagenhjelm and Agrell [12] studied a special form of index assignment based on using properties of the Hadamard transform.

These prior methods have usually been based on assuming a binary symmetric channel and/or a quantiser with a mean-square-error distortion measure, and analyses of their performance, if any, have been based on strict assumptions about the channel error or the index assignment itself. A natural approach is to formulate the index assignment in terms of an optimisation problem. In spite of its importance, there has not been any prior rigorous analysis of this problem using an optimisation formulation. Farvardin [3] used simulated annealing to find a solution for this mapping for the special case of a mean-square distortion measure. The formulation and the approach followed in [3], however, heavily depends on the assumption of a mean-square-error distortion measure.

In the present work, this mapping is formulated in terms of a quadratic programming (QP) problem. The proposed formulation optimises the combined effects of the source/channel coding on the end-to-end distortion. The formulation is general in the sense that it can handle any discrete channel model and any additive distortion measure. It will be shown that this formulation will result in an non-convex QP problem. Citing the fact that computing that exact solution to a non-convex QP problem is equivalent to solving an NP-completeness problem, we will show how an efficient interior-point algorithm can be used to find local optimum solutions which satisfy nontrivial bounds in terms of system performance.

After an explanation of the block diagram of the system, the optimisation problem under consideration is first expressed in terms of a zero-one program. It is then shown that, due to the special structure of the problem, one can relax the zero-one constraint without affecting the solution. This results in a QP problem whose solution is substantially simpler to compute than the original zero-one program. Some simulation results based on a text case for transmitting the output of a Linde–Buzo–Grauy quantiser (designed for a Gaussian source) through a binary symmetric channel are also given.

IEE Proceedings online no. 20041038

doi:10.1049/ip-com:20041038

Paper first received 3rd December 2003 and in revised form 22nd July 2004

A. Miri is with the School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

E. Hons and A.K. Khandani are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

E-mail: khandani@shannon.uwaterloo.ca

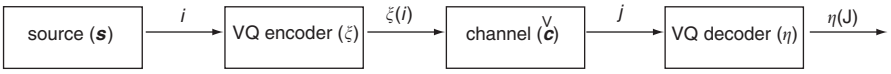

Fig. 1 *Block diagram of a typical vector quantised communication system*

2 System block diagram

Consider the communication system shown in Fig. 1. The system is composed of a source $\boldsymbol{S}$, a channel $\boldsymbol{C}$, an encoder ξ and a decoder η. The source $\boldsymbol{S}$ is composed of N_s symbols s_i, $i=0,\ldots, N_s-1$. The symbol $s_i \in \boldsymbol{S}$ occurs with probability $P_s(i)$. A measure of distortion is defined between each pair of the source symbols. The distortion between the symbols s_i, $s_k \in \boldsymbol{S}$ is denoted as $D_s(i, k)$, $i, k=0, \ldots, N_s-1$. The channel $\boldsymbol{C}$ is composed of N_c symbols c_j, $j=0, \ldots, N_c-1$. The transition probabilities of the channel, namely the probability of receiving symbol c_j when symbol c_i is transmitted, are denoted as $T_c(j|i)$, $i, j=0, \ldots, N_c-1$.

The encoder provides a mapping, denoted as ξ, from the set of source symbols to the set of channel symbols, such that the ith source symbol, $i=0, \ldots, N_s-1$, is mapped to the channel symbol indexed by $\xi(i) \in [0, N_c-1]$. Each source symbol is encoded to a specific channel symbol. However, several source symbols may be encoded to the same channel symbol, and some of the channel symbols may not be used.

The decoder provides a mapping, denoted as η, from the set of channel symbols to the set of source symbols, such that the jth channel symbol, $j=0, \ldots, N_c-1$, is mapped to the source symbol indexed by $\eta(j) \in [0, N_s-1]$. Each channel symbol is decoded to a specific source symbol, however, several channel symbols may be decoded to the same source symbol.

Our objective is to optimise the mappings ξ, η by minimising the average distortion between the encoder input and the decoder output. In the following, this optimisation problem is formulated as a zero-one program. The symbols used in this paper are listed in Table 1.

Table 1: Summary of symbols used

Symbol	Description
$P_s(i)$	occurance probability of a symbol source s_i
ξ	VQ encoder
η	VQ decoder
$D_s(i, k)$	distortion between the source symbols s_i and s_k
$T_c(j\|i)$	transition probability of receiving channel symbol c_j if c_i was transmitted
N_s	number of source symbols
N_c	number of channel symbols

3 Zero-one programming formulation

We assign an N_c dimensional binary vector to each symbol of the source at the channel input. The vector corresponding to the ith source symbol, $i=0, \ldots, N_s-1$, is denoted as $\boldsymbol{e}_i = [e_i(j), j=0, \ldots, N_c-1]$. We impose the constraints that $e_i(j) \in \{0, 1\}$, $\sum_{j=0}^{N_c-1} e_i(j) = 1$, $\forall i$. If the ith source symbol is encoded to the αth channel symbol, we set $e_i(j)=1$, $j=\alpha$ and $e_i(j)=0$, $j \neq \alpha$. We assign an N_s dimensional binary vector $\boldsymbol{d}_l = [d_l(k), k=0, \ldots, N_s-1]$, $l=0, \ldots, N_c-1$ to each channel symbol at the decoder side. If the lth channel symbol is decoded to the βth source symbol, we set $d_l(k)=1$, $k=\beta$ and $d_l(k)=0$, $k \neq \beta$.

Using these notations, the optimisation problem is formulated as:

$$\text{minimise} \sum_{i=0}^{N_s-1} \sum_{j=0}^{N_c-1} \sum_{l=0}^{N_c-1} \sum_{k=0}^{N_s-1} P_s(i) T_c(l|j) D_s(i,k) e_i(j) d_l(k) \tag{1}$$

subject to:

$$e_i(j) \in \{0, 1\} \quad \text{and} \quad \sum_{j=0}^{N_c-1} e_i(j) = 1, \forall i \tag{2}$$

$$d_l(k) \in \{0, 1\} \quad \text{and} \quad \sum_{k=0}^{N_s-1} d_l(k) = 1, \forall l \tag{3}$$

We refer to the constraints $\sum_j e_i(j) = 1$ and $\sum_k d_l(k) = 1$ as the 'indicator constraints'. Note that the indicator constraints are non-overlapping, in the sense that each of them involves a different set of variables. Due to this structure, the extreme points of the polytopes corresponding to $\sum_j e_i(j) = 1$ and $\sum_k d_l(k) = 1$ are composed of zero-one variables only.

The introduced formulation optimises the combined effects of source quantisation and channel coding on the end-to-end distortion. It is well known that providing a proper trade-off between source and channel coding plays an important role in the performance of a digital communication systems [13]. We note that quantisation of a set of source symbols is equivalent to a grouping of those symbols into disjoint partitions and using a single representative for each partition. In this sense, quantisation of the source symbols occurs when several source symbols are encoded to the same channel symbol. We also note that channel coding is equivalent to discarding some of the symbols at the channel input to increase the level of the immunity against the channel errors. As already mentioned, this possibility is incorporated in our coding structure.

The immediate problem in applying optimisation methods to solve (1) is that the variables are restricted to 0 and 1 (zero-one program). It is generally known that solving a zero-one program is a very hard problem. However, one can relax the zero-one constraint as is explained in the following.

The constraints of

$$e_i(j) \in \{0, 1\}, \quad \sum_{j=0}^{N_c-1} e_i(j) = 1$$

reflect the restriction that each input source symbol is mapped to a unique channel symbol. Similarly, the constraints of

$$d_l(k) \in \{0, 1\}, \quad \sum_{k=0}^{N_s-1} d_l(k) = 1$$

reflect the restriction that each output channel symbol is mapped to a unique source symbol. We refer to a construction with such constraints as deterministic. In an

alternative, we replace these constraints by

$$e_i(j) \geq 0, \sum_{j=0}^{N_c-1} e_i(j) = 1$$

and

$$d_l(k) \geq 0, \sum_{k=0}^{N_s-1} d_l(k) = 1$$

respectively. In this case, the resulting encoder and the decoder mappings will be stochastic, in the sense that the ith source symbol is mapped with probability $e_i(j)$ to the jth input channel symbol, and the lth output channel symbol is mapped with probability $d_l(k)$ to the kth source symbol.

As the optimisation problem resulting in a stochastic structure is obtained by a relaxation of the corresponding problem in the deterministic case, such a stochastic mapping cannot result in a degradation in performance. Indeed, noting that the objective function in (1) is bilinear in terms of $e_i(j)$ and $d_l(k)$, it is easy to see that the final solution of the relaxed problem will be located at an extreme point of the polytope corresponding to the $e_i(j)$ constraints (linearity with respect to $e_i(j)$ s), as well as at an extreme point of the polytope corresponding to the $d_l(k)$ constraints (linearity with respect to $d_l(k)$ s). As a result, $e_i(j)$ and $d_l(k)$ variables will be composed of zero-one values only, i.e. the zero-one constraint is automatically satisfied. In the following, the optimisation problem in (1) is transformed into a QP problem.

4 Quadratic programming problem

The quadratic programming problem used here may be stated as:

$$\begin{array}{lll} \text{minimise} & q(\boldsymbol{x}) & = \frac{1}{2}\boldsymbol{x}^t \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to:} & \boldsymbol{x} \in \boldsymbol{X} & = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq 0\} \end{array} \quad (4)$$

where $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ for some positive integers m and n.

The problem in (1) can be expressed in terms of a QP problem by arranging the elements $e_i(j)$ and $d_l(k)$ in the column vector $\boldsymbol{x}$, i.e.

$$\begin{aligned} \boldsymbol{x}^t = [&e_0(0), e_0(1), \ldots, e_0(N_c-1), \ldots, e_{N_s-1}(0), \ldots, \\ &e_{N_s-1}(N_c-1), \ldots, d_0(0), \ldots, d_{N_c-1}(N_s-1)] \end{aligned}$$

and

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{D} \\ \boldsymbol{D}^t & \boldsymbol{0} \end{bmatrix}$$

where $\boldsymbol{D}$ is an $N_sN_c \times N_sN_c$ square matrix with $P_s(i)T_c(l|j)$ entries corresponding to entries of the vector $\boldsymbol{x}$ in (1). For example, consider the case where $N_s = 2$ and $N_c = 1$. Then, for

$$\boldsymbol{x}^t = [e_0(0), e_0(1), e_1(0), e_1(1), d_0(0), d_0(1), d_1(0), d_1(1)]$$

the corresponding matrix $\boldsymbol{D}$ is given by

$$\begin{bmatrix} P_s(0)T_c(0|0)D_s(0,0) & P_s(0)T_c(0|0)D_s(0,1) \\ P_s(0)T_c(0|1)D_s(0,0) & P_s(0)T_c(0|1)D_s(0,1) \\ P_s(1)T_c(0|0)D_s(1,0) & P_s(1)T_c(0|0)D_s(1,1) \\ P_s(1)T_c(0|1)D_s(1,0) & P_s(1)T_c(0|1)D_s(1,1) \end{bmatrix}$$
$$\begin{bmatrix} P_s(0)T_c(1|0)D_s(0,0) & P_s(0)T_c(1|0)D_s(0,1) \\ P_s(0)T_c(1|1)D_s(0,0) & P_s(0)T_c(1|0)D_s(0,1) \\ P_s(1)T_c(1|0)D_s(1,0) & P_s(1)T_c(1|0)D_s(1,1) \\ P_s(1)T_c(1|1)D_s(1,0) & P_s(1)T_c(1|1)D_s(1,1) \end{bmatrix}$$

The indicator constraints described by (2) and (3) can be also rewritten as the constraints on the QP problem (4) by simply letting

$$\boldsymbol{A} = \begin{bmatrix} 1\cdots1 & & \\ & \ddots & \\ & & 1\cdots1 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \Bigg\} N_c + N_s$$

It is well known that the QP problem (4) is a non-convex problem if the matrix $\boldsymbol{Q}$ is not positive semi-definite. In the case dealt with in the paper, the matrix $\boldsymbol{Q}$ is not positive semi-definite as will be shown below.

To show that the matrix $\boldsymbol{Q}$ in (4) is not positive semi-definite, recall the structure of $\boldsymbol{Q}$ and the fact that the matrix $\boldsymbol{D}$ is a square matrix. It is easy to see that the main diagonal elements of the $\boldsymbol{Q}$ matrix are always exactly zero. To test if the $\boldsymbol{Q}$ matrix is positive semi-definite, we use the following lemma.

Lemma: [14] If a symmetric matrix $\boldsymbol{Q} = (q_{i,j})$ is positive semi-definite and has a zero diagonal entry $q_{i,i}$, then all the entries in that row and column have to be zero, i.e., the row i and column i must be zero.

The lemma above and the fact that the main diagonal entries of the $\boldsymbol{Q}$ matrix are always zero, implies that the only case where the $\boldsymbol{Q}$ matrix is positive semi-definite is when all its entries are equal to zero, making the objective function identical to zero.

It has long been known that a non-convex QP problem is a hard problem to solve. This was first shown by Sahni [15], and later by Pardalos [16]. More recently, several authors have shown that a non-convex QP is actually NP-complete [17]. This will imply that a polynomial algorithm for computing the exact solution of this problem cannot be expected, since that would imply $P = NP$.

Having shown that the optimisation problem stated in (1) is a non-convex QP, the approach will be to find a local optimum solution which satisfies some nontrivial bound in terms of how well it minimises the objective function.

A feasible point x is defined to be an ε-approximation if there exists a global minimum x^* such that $q(x) - q(x^*) \leq \varepsilon$ for some $\varepsilon > 0$. However, we will use the following definition since some of its useful properties prove crucial in our approach. We assume that the feasible region is bounded and closed, and contains a nonempty interior region.

Definition: Assume that an upper bound for $P_s(i)T_c(l|j)D_s(i, k)$, or in the case where the probabilities of source symbol $P_s(i)$ and the transition probabilities are known an upper bound for distortion $D_s(i, k)$ is known. Let $\bar{z}$ be the value of the objective function in (4) if all $P_s(i)T_c(l|j)D_s(i, k)$, or $D(i, k)$ are replaced by this upper bound. Also, let $\underline{z}$ denote where the objective function is minimised. A feasible point $\boldsymbol{x}$ is an ε-minimal solution or ε-approximation, $\varepsilon \in (0, 1)$ for (4) if

$$\frac{q(x) - \underline{z}}{\bar{z} - \underline{z}} \leq \varepsilon$$

This definition has the advantage that the ε-approximation property of a point is preserved under affine linear transformations of the feasible region. Note that only an optimum point is a zero-approximation.

Another useful property of the above definition is that it is insensitive to translation or dilation of the objective function. In other words, let the objective function $q(\boldsymbol{x})$ be replaced by a new objective function $f(\boldsymbol{x}) = aq(\boldsymbol{x}) + b$ for some $a > 0$. It is easy to see that an ε-approximate point $\boldsymbol{x}$ will be also an ε-approximate of the new objective function.

Our approach to solving the QP problem (4) is to utilise the fact that efficient algorithms exist for certain non-convex

quadratic minimisation problems (in particular minimisation is subject to a unit ball constraint i.e. $\|x\|_2 \leq 1$). Interior-point algorithms for linear programming often use a gradient direction method. However, this will give a substantial reduction in the objective function only if the current feasible point is centred in a somewhat large polytope. Therefore a commonly used technique is to use a scheme that alternates between centring the feasible point and taking a step in the gradient direction. The interior-point algorithm described here uses the affine scaling algorithm, first proposed by Dikin [18], followed by a polynomial time bisection algorithm that solves a sub-optimisation ball-constraint QP problem. This will produce an ε-approximation to a local minimum associated with (4) which satisfies the Karush–Kuhn–Tucker (KKT) first and second conditions.

In order to solve the QP problem (4), given any arbitrary feasible point $\boldsymbol{x}^k \in \boldsymbol{X}$, where $\boldsymbol{x}^k = (x_1^k, \ldots, x_n^k) > 0$, we use affine scaling techniques to minimise the objective function over an interior ellipsoid which will generate another interior solution. A series of such ellipsoids can be successively constructed to generate a sequence of points converging to a KKT point [19, 20]. Note that in the above, k refers to the kth iteration of the affine scaling technique, and it is assumed that $\boldsymbol{x}^0$ is given. Let $\boldsymbol{X}^k$ be the $n \times n$ diagonal matrix known as the scaling matrix, whose jth diagonal entry is x_j^k. If one or more x_j^k are zero, a perturbed diagonal scaling matrix is used for which the zero elements are replaced by, for example, a fixed but non-zero element. Then let $\boldsymbol{y} = \boldsymbol{X}^k \boldsymbol{x}$, $\boldsymbol{Q}^k = \boldsymbol{X}^k \boldsymbol{Q} \boldsymbol{X}^k$ and $\boldsymbol{A}^k = \boldsymbol{A}\boldsymbol{X}^k$. Affine scaling solves the following sub-optimisation problem:

$$\begin{array}{ll} \text{minimise} & q(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^t \boldsymbol{Q}^k \boldsymbol{y} \\ \text{subject to:} & \boldsymbol{y} \in \{\boldsymbol{y} \in \mathbb{R}^n : A^k \boldsymbol{y} = b,\ \|\boldsymbol{y} - \mathcal{E}\| \leq \alpha\} \end{array} \tag{5}$$

where $\|\cdot\|$ denotes the L_2-norm, $0 < \alpha < 1$, $\mathcal{E}^t = (1, \ldots, 1)$, and $\{\boldsymbol{x} : \left\|(X^k)^{-1}(\boldsymbol{x} - \boldsymbol{x}^k)\right\| = \|\boldsymbol{y} - \mathcal{E}\| \leq \alpha\}$ corresponds to the Dikin ellipsoid with radius α in the positive octant $\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > 0\}$. Note that the ellipsoid constraint $\{\|\boldsymbol{y} - \varepsilon\| \leq \alpha < 1\}$ implies that any feasible solution $\boldsymbol{y}$ of (5) gives a positive interior feasible solution $\boldsymbol{x} = (X^k)^{-1}\boldsymbol{y}$ for (4).

There are several approximation algorithms developed for QP which minimise the objective function in an L_2-norm neighbourhood around a given feasible point (a ball constraint). The basic idea is to use an affine transformation of the ellipsoid constraint QP which can be solved using a ball constraint QP [20].

Let $\boldsymbol{N}^k \in \mathbb{R}^{n \times (n-m)}$ be a matrix whose columns form an orthonormal basis spanning the null space of $\boldsymbol{A}^k$, and let $\boldsymbol{N}^k \boldsymbol{z} = \boldsymbol{y} - \mathcal{E}$ for some $\boldsymbol{z} \in \mathbb{R}^{n-m}$. Also, let $\boldsymbol{H}^k = (\boldsymbol{N}^k)^t \boldsymbol{Q}^k \boldsymbol{N}^k \in \mathbb{R}^{(n-m) \times (n-m)}$ and $\boldsymbol{g}^k = (\boldsymbol{Q}^k \mathcal{E})^t \boldsymbol{N}^k \in \mathbb{R}^{n-m}$. Then, the QP problem, (5) can be written as,

$$\begin{array}{ll} \text{minimise} & q(\boldsymbol{z}) = (\boldsymbol{g}^k)^t \boldsymbol{z} + \frac{1}{2}\boldsymbol{z}^t \boldsymbol{H}^k \boldsymbol{z} \\ \text{subject to:} & \boldsymbol{z} \in \{\boldsymbol{z} \in \mathbb{R}^{n-m} : \|\boldsymbol{z}\| \leq \alpha\} \end{array} \tag{6}$$

To compute the ε-approximation of (6), we use a simple polynomial bisection method proposed by Fu *et al.* [21] (see also [20, 22]). This results in a total of

$$O\left(\left(\frac{(n-m)^6}{\varepsilon}\log\frac{1}{\varepsilon} + (n-m)^4 \log(n-m)\right) \times \left(\log\frac{1}{\varepsilon} + \log(n-m)\right)\right) \tag{7}$$

arithmetic operations. For the details of how (6) can be obtained from (5), and a sketch of the algorithm to solve the ball constraint QP problem using a bisection method, please see the Appendix.

Let $\underline{\boldsymbol{z}}$ be the minimal solution of (6). Generate a sequence of points by letting $\underline{\boldsymbol{z}}^{k+1} = X^k \underline{\boldsymbol{z}}$. It is proved in [20] that these points converge to an optimal point satisfying both the first and second necessary KKT conditions for the original QP problem (4).

5 Simulation results

The test cases simulate transmission of a quantised Gaussian source. The Linde–Buzo–Gray (LBG) algorithm is used to produce the quantiser points, which become the communication systems' source symbols. We use the same number of source and channel symbols. A binary symmetric channel (BSC) model is used within which a binary symbol is received in error with a given probability, say ε. Multiple channel symbols are obtained by simulating repeated use of the BSC. In this case, we have for the channel transfer matrix T the following values:

$$T_c(j|i) = (\varepsilon)^{h_{i,j}}(1-\varepsilon)^{(r-h_{i,j})} \tag{8}$$

where r is the binary length of the code words and $h_{i,j}$ is the Hamming distance (number of bit differences) between code words indexed by i and j.

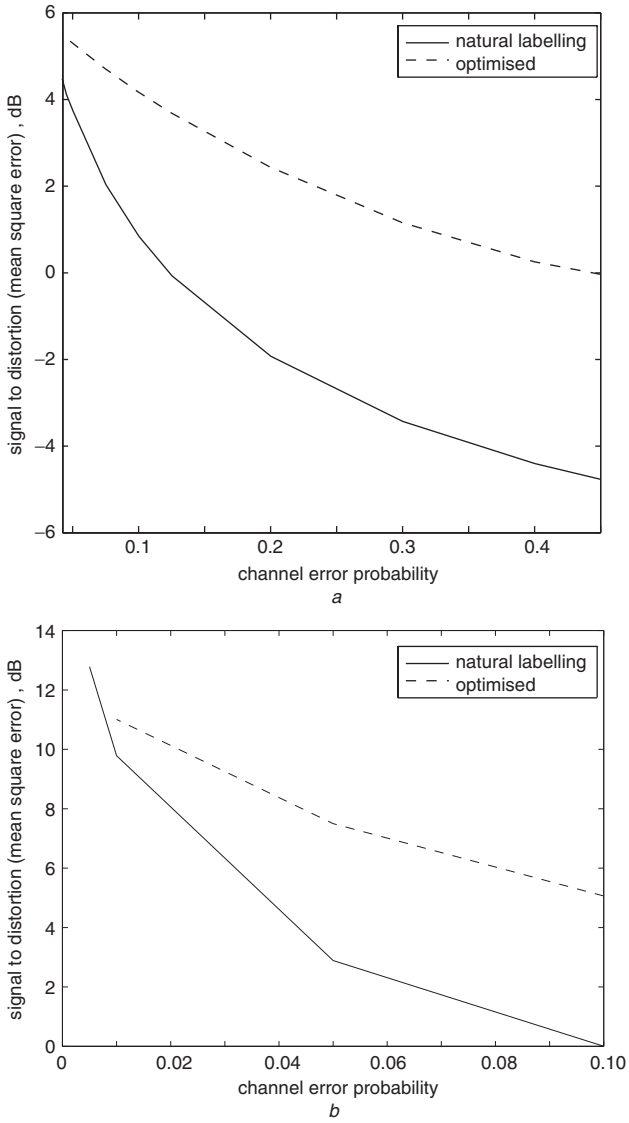

Fig. 2 *Natural labelling vs optimised encoder/decoder performance*
a 8 symbols
b 16 symbols

The graphs, Fig. 2, show the ratio of discrete source energy to added distortion (in dB) plotted against the channel error parameter (ε). The energy of a discrete source $S = s_1, \ldots, s_n$ with a priori probabilities $p_1, \ldots, p_n$ is given by $E = \sum_i p_i s_i^2 - (\sum_i p_i s_i)^2$. The added distortion is the value of the objective function from (1) or any of the later reformulations. In both graphs, the optimised mapping is compared with a natural labelling which maps the source symbol s_i to the channel symbol corresponding to the binary representation of i and has the inverse mapping as the decoder. This corresponds to a typical system without channel coding.

It is observed that the optimisation procedure explained in this paper can result in a substantial improvement in the end to end distortion of the source, especially for poor channel conditions.

6 Conclusions and future work

The problem of optimising the mapping between the source and the channel symbols to minimise the average distortion between the encoder input and the decoder output of a discrete communication system is addressed. It is shown how this problem can be formulated in terms of a zero-one optimisation problem. This combined source-channel coding formulation allows for the case when several source symbols are encoded to the same channel symbol, as well as for the case when some of the channel symbols are not used at all, resulting in a larger noise margin for the remaining symbols. It can also handle any discrete channel model and any additive distortion measure. Using the structure of the problem, it is shown that the zero-one formulation can be reduced to a general QP problem. Affine scaling techniques, followed by a bisection algorithm which solved a sub-problem in a polynomial time bound of $O(n^6L)$ was used to generate an ε-approximate solution of the given QP problem. Numerical results are presented for some cases of practical interest in digital communications, demonstrating a substantial improvement using the proposed method.

In this paper, the index assignment problem was reformulated as a QP problem and a possible solution based on an interior point method was suggested. The study of constrained QP problems has been one of the richest areas in optimisation theory. Other techniques used to solve this problem could be used in future work, and comparisons made in terms of complexity and convergence rate.

As mentioned in the Introduction, there have been many methods developed in order to improve the system performance, in the presence of channel noise, without using explicit error control. A comprehensive survey of these methods and their advantages, disadvantages and restrictions, and the method suggested in this paper could provide a very useful tool in this area.

In this paper, we did not address the complexity of the implementation of the vector quantisation encoder/decoder where the number of source symbols is large. An adaptation of a hierarchical lookup table, similar to the one suggested in [23], could potentially offer a solution to this problem.

7 Acknowledgments

The authors would like to express their thanks to the anonymous reviewers for their constructive suggestions that have helped the authors to improve the presentation and readability of this paper.

This work is financially supported by and Natural Sciences and Engineering Research Council of Canada (NSERC) and by Communications and Information Technology Ontario (CITO).

8 References

1 Chen, J.H., Davidson, G., Gersho, A., and Zeger, K.: 'Speech coding for the mobile satellite experiment'. Proc. IEEE Int. Conf. on Communications, Seattle, WA, USA, 1987, pp. 756–763
2 DeMarca, J.R.B., and Jayant, N.S.: 'An algorithm for assigning binary indices to the code vectors of a multi-dimensional quantizer'. Proc. IEEE Int. Conf. on Communications, Seattle, WA, USA, 1987, pp. 1128–1132
3 Farvardin, N.: 'A study of vector quantization for noisy channels', *IEEE Trans. Inf. Theory*, 1990, **36**, pp. 799–809
4 Zeger, K.A., and Gersho, A.: 'Zero redundancy channel coding in vector quantization', *Electron. Lett.*, 1987, **23**, pp. 654–655
5 Zeger, K., and Gersho, A.: 'Pseudo-Gray coding', *IEEE Trans. Commun.*, 1990, **38**, (12), pp. 2147–2156
6 Ben-David, G., and Malah, D.: 'On the performance of a vector quantizer under channel errors', in Vandewalle, J., Boite, R., Moonen, M. and Oosterlinck, A. (Eds.): 'Signal processing VI: theories and applications, Proceeding of EUSIPCO'92' (Elsevier Science Publishers, 1992), pp. 1685–1688
7 Hagen, R., and Hedelin, P.: 'Robust vector quantization by linear mappings of block-codes'. Proc. IEEE Int. Symp. Information Theory, San Antonio, TX, USA, 1993, p. 171
8 Hagen, R., and Hedelin, P.: 'Design methods for VQ by linear mappings of block codes'. Proc. IEEE Int. Symp. on Information Theory, Trondheim, Norway, 1994, p. 241
9 McLaughlin, S.W., Neuhoff, D.L., and Ashley, J.K.: 'Optimal binary index assignments for a class of equiprobable scalar and vector quantizers', *IEEE Trans. Inf. Theory*, 1995, **41**, (6), pp. 2031–2037
10 McLaughlin, S.W., and Neuhoff, D.L.: 'Neuhoff Asymptotic bounds in source channel coding'. Proc. IEEE Int. Symp. on Information Theory, Budapest, Hungary, 1991, p. 61
11 Zeger, K., and Manzella, V.: 'Asymptotic bounds on optimal noisy channel quantization via randm coding', *IEEE Trans. Inf. Theory*, 1994, **40**, (6), pp. 1926–1938
12 Knagenhjelm, P., and Agrell, E.: 'The Hadamard transform-A tool for index assignment', *IEEE Trans. Inf. Theory*, 1996, **42**, pp. 1139–1151
13 Hochwalk, B., and Zegger, K.: 'Tradeoff between source and channel coding', *IEEE Trans. Inf. Theory*, 1997, **43**, (5), pp. 1412–1424
14 Murty, K.G.: 'Operations research: deterministic optimization models' (Prentice-Hall Inc, 1995)
15 Sahni, S.: 'Computationally related problems', *SIAM J. Comput.*, 1974, **3**, pp. 303–324
16 Pardalos, P.M.: 'Global optimization algorithms for linearly constrained indefinite quadratic problems', *Comput. Math. Appl.*, 1991, **21**, pp. 87–97
17 Vavasis, S.A.: 'Quadratic programming is in NP', *Inf. Process. Lett.*, 1990, **36**, pp. 73–77
18 Dikin, I.I.: 'Iterative solution of problems of linear and quadratic programming', *Sov. Math. Dokl.*, 1967, **8**, pp. 674–675
19 Ye, Y.: 'An $O(n^3L)$ potential reduction algorithm for linear programming', *Math. Program.*, 1991, **50**, pp. 239–258
20 Ye, Y.: 'On affine scaling algorithm for non-convex quadratic programming', *Math. Program*, 1992, **56**, pp. 285–300
21 Fu, M., Luo, Z.-Q., Ye, Y.: 'Approximation algorithms for quadratic programming'. Manuscript, Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada, 1996
22 Vavasis, S.A.: 'Polynomial time weak approximation algorithms for quadratic programming', in Floudas, C.A. and Pardalos, P.M. (Eds.): 'Complexity in numerical optimization' (World Scientific, 1993), pp. 490–500
23 Jafarkhani, H., and Farvardin, N.: 'Design of channel-optimized vector quantizer in the presence of channel mismatch', *IEEE Trans. Commun.*, 2004, **48**, (1), pp. 118–124
24 Ye, Y.: 'Interior point algorithms: theory and analysis', Wiley-interscience series in discrete mathematics and optimization (John Wiley and Sons, 1997)

9 Appendix

This Appendix shows the relationships between different QPs presented in Section 4, and sketches of the algorithm which solves the ball constraint QP problem using a bisection method.

Consider the general case of the QP problem (5):

$$\begin{array}{lll} \text{minimise} & q(\boldsymbol{x}) & = \frac{1}{2}\boldsymbol{x}^t\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to:} & & \boldsymbol{x} \in \{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = b,\ \boldsymbol{x} \geq 0\} \end{array} \tag{9}$$

Consider an arbitrary feasible point $\boldsymbol{x}^k \in \boldsymbol{X}$, where $\boldsymbol{x}^k = (x_1^k, \ldots, x_n^k) > 0$. Let $\boldsymbol{X}^k = \text{diag}[x_1^k, \ldots, x_n^k]$ be the diagonal scaling matrix, and use a change variable $\boldsymbol{y} = \boldsymbol{X}^k\boldsymbol{x}$.

Setting $\boldsymbol{Q}^k = \boldsymbol{X}^k \boldsymbol{Q} \boldsymbol{X}^k$ and $\boldsymbol{A}^k = \boldsymbol{A}\boldsymbol{X}^k$, and for $\mathcal{E}^t = (1, \ldots, 1)$ we can then solve the following sub-optimisation problem:

$$\begin{aligned} &\text{minimise} \quad q(\boldsymbol{y}) \quad = \tfrac{1}{2}\boldsymbol{y}^t \boldsymbol{Q}^k \boldsymbol{y} \\ &\text{subject to :} \quad \boldsymbol{y} \in \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{A}^k \boldsymbol{y} = \boldsymbol{b},\ \|\boldsymbol{y} - \mathcal{E}\| \le \alpha\} \end{aligned} \tag{10}$$

It should be noted that since $\boldsymbol{x}^k$ is a feasible point, we have

$$\boldsymbol{A}^k \mathcal{E} = \boldsymbol{A}\boldsymbol{X}^k \mathcal{E} = \boldsymbol{A}\boldsymbol{x}^k = b$$

We also have that matrix $\boldsymbol{Q}^k$ is symmetric, since $\boldsymbol{Q}$ is. In fact, $\boldsymbol{Q}^k$ is an $n \times n$ diagonal matrix whose jth diagonal element is $(x_k^k)^t Q(x_j^k)$.

Now let $\Delta \boldsymbol{y} = \boldsymbol{y} - \mathcal{E}$. It is easy to show that the following set of equalities hold:

$$\begin{aligned} \mathcal{E}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} &= \Delta \boldsymbol{y}^t \boldsymbol{Q}^k \mathcal{E} = \mathcal{E}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} = \mathcal{E}^t (\boldsymbol{Q}^k)^t \Delta \boldsymbol{y} \\ &= (\boldsymbol{Q}^k \mathcal{E})^t \Delta \boldsymbol{y} \end{aligned} \tag{11}$$

The last three equalities follow from the fact that $\boldsymbol{Q}^k$ is symmetric, and by a basic property of transpositions. To show the first two equalities, note that

$$\begin{aligned} &\mathcal{E}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} = (1 \quad 1 \quad \cdots \quad 1) \\ &\begin{pmatrix} (x_1^k)^t Q(x_1^k) & & & \\ & (x_2^k)^t Q(x_2^k) & & \\ & & \ddots & \\ & & & (x_n^k)^t Q(x_n^k) \end{pmatrix} \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_n \end{pmatrix} \\ &= (x_1^k)^t Q(x_1^k) \Delta \boldsymbol{y}_1 + (x_2^k)^t Q(x_2^k) \Delta \boldsymbol{y}_2 + \cdots + (x_n^k)^t Q(x_n^k) \Delta \boldsymbol{y}_n \end{aligned}$$

which is the same as

$$\begin{aligned} &\Delta \boldsymbol{y}^t \boldsymbol{Q}^k \mathcal{E} = (\Delta \boldsymbol{y}_1 \quad \Delta \boldsymbol{y}_2 \quad \cdots \quad \Delta \boldsymbol{y}_n) \\ &\begin{pmatrix} (x_1^k)^t Q(x_1^k) & & & \\ & (x_2^k)^t Q(x_2^k) & & \\ & & \ddots & \\ & & & (x_n^k)^t Q(x_n^k) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \end{aligned}$$

Now using the equalities in (11) and the fact that $\boldsymbol{y} = \Delta \boldsymbol{y} + \mathcal{E}$, $q(\boldsymbol{y})$ can be written as

$$\begin{aligned} q(\boldsymbol{y}) &= \frac{1}{2}(\Delta \boldsymbol{y} + \mathcal{E})^t \boldsymbol{Q}^k (\Delta \boldsymbol{y} + \mathcal{E}) \\ &= \frac{1}{2}(\Delta \boldsymbol{y}^t + \mathcal{E}^t) \boldsymbol{Q}^k (\Delta \boldsymbol{y} + \mathcal{E}) \\ &= \frac{1}{2}(\Delta \boldsymbol{y}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} + \left(\frac{1}{2}\Delta \boldsymbol{y}^t \boldsymbol{Q}^k \mathcal{E} + \frac{1}{2}\mathcal{E}^t \boldsymbol{Q}^k \Delta \boldsymbol{y}\right) + \frac{1}{2}\mathcal{E}^t \boldsymbol{Q}^k \mathcal{E} \\ &= \frac{1}{2}(\Delta \boldsymbol{y}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} + (\boldsymbol{Q}^k \mathcal{E})^t \Delta \boldsymbol{y} + \frac{1}{2}\mathcal{E}^t \boldsymbol{Q}^k \mathcal{E} \end{aligned}$$

Since $q(\mathcal{E}) = \frac{1}{2}\mathcal{E}^t \boldsymbol{Q}^k \mathcal{E}$ is a constant and hence can be ignored in the minimisation problem, and given the fact that we have already shown

$$\boldsymbol{A}^k(\Delta \boldsymbol{y} + \mathcal{E}) = \boldsymbol{A}^k \Delta \boldsymbol{y} + \boldsymbol{A}^k \mathcal{E} = \boldsymbol{A}^k \Delta \boldsymbol{y} + \boldsymbol{b}$$

the QP (10) can be rewritten as

$$\begin{aligned} &\text{minimise} \quad \tfrac{1}{2}\Delta \boldsymbol{y}^t \boldsymbol{Q}^k \Delta \boldsymbol{y} + (\boldsymbol{Q}^k \mathcal{E})^t \Delta \boldsymbol{y} \\ &\text{subject to:} \quad \boldsymbol{y} \in \{\Delta \boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{A}^k \Delta \boldsymbol{y} = 0,\ \|\Delta \boldsymbol{y}\| \le \alpha\} \end{aligned} \tag{12}$$

Now, let $\boldsymbol{N}^k \in \mathbb{R}^{n \times (n-m)}$ be a matrix whose columns form an orthonormal basis spanning the null space of $\boldsymbol{A}^k$, such that $(\boldsymbol{N}^k)^t(\boldsymbol{N}^k) = \boldsymbol{I}$, and let $\boldsymbol{N}^k \boldsymbol{z} = \Delta \boldsymbol{y}$. Then, the QP problem (12) can be written as the ball-constraint QP,

$$\begin{aligned} &\text{minimise} \quad q(\boldsymbol{z}) \quad = \tfrac{1}{2}\boldsymbol{z}^t \boldsymbol{H}^k \boldsymbol{z} (\boldsymbol{g}^k)^t \boldsymbol{z} \\ &\text{subject to:} \quad \boldsymbol{z} \in \{\boldsymbol{z} \in \mathbb{R}^{n-m} : \|\boldsymbol{z}\| \le \alpha\} \end{aligned} \tag{13}$$

where $\boldsymbol{H}^k = (\boldsymbol{N}^k)^t \boldsymbol{Q}^k \boldsymbol{N}^k \in \mathbb{R}^{(n-m) \times (n-m)}$ and $\boldsymbol{g}^k = (\boldsymbol{Q}^k \mathcal{E})^t \boldsymbol{N}^k \in \mathbb{R}^{n-m}$. Here, we have used the fact that $\|\boldsymbol{N}^k \boldsymbol{z}\| = \|\boldsymbol{N}^k\| \|\boldsymbol{z}\| = \|\boldsymbol{z}\|$.

The ball-constraint QP (13) is then successively solved for each iteration of the affine scaling step until a local optimum point is found.

Below is the sketch of steps involved in solving (13) using a bisection method. It should be noted that the ball-constraint QP is one of the most studied problems, and the bisection method is one of the many suggested algorithms to solve it. For more details on the bisection method below, the reader is referred to [20, 24].

It is well known that the solution, $\boldsymbol{z}$, of the ball constraint QP (13) satisfies the following necessary and sufficient conditions,

$$(\boldsymbol{H}^k + \mu \boldsymbol{I})\boldsymbol{z} = -\boldsymbol{g}^k \tag{14}$$

$$\mu \ge \max\{0, -\underline{\lambda}\} \tag{15}$$

$$\text{and } \|\boldsymbol{z}\| \le \alpha \tag{16}$$

where $\underline{\lambda}$ denotes the least eigenvalue of the matrix $\boldsymbol{H}^k$. In the case where $\boldsymbol{H}^k$ is not positive semidefinite, we must have $\underline{\lambda} < 0$. Furthermore, it is known that [20]

$$|\underline{\lambda}| \le (n - m) \max \left| h_{ij}^k \right| \tag{17}$$

and

$$\mu^* \le |\underline{\lambda}| + \frac{\|g^k\|}{\alpha} \tag{18}$$

where μ^* is the unique solution to (14), and $h_{i,j}^k$ is the (i, j)th component of the matrix $\boldsymbol{H}^k$. We have,

$$0 \le \mu^* \le \mu^0 := (n - m) \max\{|h_{ij}^k|\} + \frac{\|g^k\|}{\alpha}$$

For any given $\mu > |\underline{\lambda}|$, let $\boldsymbol{z}_\mu$ be a solution to the linear equation (14). We can use the bisection method stated below to look for the root of $\|\boldsymbol{z}_\mu\|$ over the interval $\mu \in [|\underline{\lambda}|, \mu^0] \subset [0, \mu^0]$. It was proved in [21] that this will generate an ε-minimiser of (13) in the polynomial time stated in (7).

Algorithm: for solving ball-constraint QP (13)

Set $\mu_1 = 0$ and $\mu_3 = n \max_{ij} |h_{ij}| + \|\boldsymbol{g}^k\|/\alpha$, and

stop: = *false*.

Step 1: Set $\mu_2 = \frac{1}{2}(\mu_1 + \mu_3)$

Step 2: Let $\mu = \mu_2$, and solve for $\boldsymbol{z}$ in (14)

Step 3: **If** $\mu_3 - \mu_1 < \varepsilon$ **then** *stop:* = *true*;

elseif

(*a*) $H^k + \mu I$ is indefinite or negative definite, or

(*b*) (14) has no solution, or

(*c*) the norm of the minimal norm solution of (14) is greater than α,

then $\mu_1 = \mu_2$ and **goto** step 1;

elseif the norm of the solution of (14) is less than α,

then $\mu_3 = \mu_2$ and **goto** step 1.