Joint Entropy-Constrained Multiterminal Quantization

Jean Cardinal
Université Libre de Bruxelles CP 212
B-1050 Brussels, Belgium
email:jcardin@ulb.ac.be

Gilles Van Assche
Université Libre de Bruxelles CP 165
B-1050 Brussels, Belgium
email:gvanassc@ulb.ac.be

Abstract — **We study the design of entropy-constrained multiterminal quantizers for coding two correlated continuous sources. Two design algorithms are presented, both optimizing a Lagrangian cost measure involving distortions and information rates.**

I. Introduction

We study the design of a multiterminal quantizer (Fig. 1) for the encoding of two correlated sources X and Y in $\mathbb{R}^k$. We assume that the sources are encoded separately by the encoders α_X and α_Y, and that the pair of output indices is jointly decoded by the decoder β. The reproduction values are denoted by $\hat{X}$ and $\hat{Y}$, and both corresponding distortions $E[d(X,\hat{X})]$ and $E[d(Y,\hat{Y})]$ must be minimized. We further assume that the quantization step is followed by an ideal Slepian-Wolf (SW, [1]) entropy coder $(\gamma_X, \gamma_Y, \gamma^{-1})$. This assumption does make sense since recent works [2] propose practical coders operating close to the SW bounds. We propose to design a quantizer pair jointly minimizing the two distortions with constraints on the two bit rates predicted by the SW theorem.

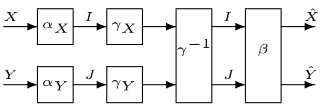

Fig. 1. Block diagram of the proposed multiterminal quantizer

II. Algorithms

We use the notations $I = \alpha_X(X)$, $J = \alpha_Y(Y)$. The achievable rates for SW coding of these two indices satisfy [1] $R_X \geq H(I \mid J)$, $R_Y \geq H(J \mid I)$ and $R_X + R_Y \geq H(I,J)$.

Using discrete Lagrangian optimization, we define the cost measure to be minimized by the triple $(\alpha_X, \alpha_Y, \beta)$: $J(\alpha_X, \alpha_Y, \beta) = D_X + \mu D_Y + \lambda_X R_X + \lambda_Y R_Y$, where D_X and D_Y are the average distortions, R_X and R_Y the average rates, and μ, λ_X and λ_Y are positive Lagrangian multipliers. We define $\beta(i,j) = (\beta_X(i,j), \beta_Y(i,j))$, and $R_X = H(I)$, $R_Y = H(J \mid I)$. This is without loss of generality, since time sharing allows to reach any point on the achievable rates curve. The optimality condition for α_X is

$$\begin{aligned}\alpha_X(x) = \quad \min_i^{-1} \quad & E_Y[d(x, \beta_X(i,J)) + \mu d(Y, \beta_Y(i,J)) \\ & -\lambda_Y \log P[J \mid I = i] \mid X = x] \\ & -\lambda_X \log P[I = i],\end{aligned}$$

for all $x \in \mathbb{R}^k$, where i is taken in the set $\mathcal{I}_X$ of indices for α_X. The optimal encoder for Y is similar:

$$\begin{aligned}\alpha_Y(y) = \quad \min_j^{-1} \quad & E_X[d(X, \beta_X(I,j)) + \mu d(y, \beta_Y(I,j)) \\ & -\lambda_Y \log P[J = j \mid I] \mid Y = y],\end{aligned}$$

for all $y \in \mathbb{R}^k$, where j is taken in the set $\mathcal{I}_Y$ of indices for α_Y. The optimal decoder for the mean squared error is the classical Bayes estimator: $\beta(i,j) = E[X, Y \mid I = i \wedge J = j]$ for all $(i,j) \in \mathcal{I}_X \times \mathcal{I}_Y$. A simple descent algorithm for the design of $(\alpha_X, \alpha_Y, \beta)$ consists in alternatively forcing the above optimality conditions. General equations of the same flavor can be found in a recent work of Fleming et al. [3]. The implementation of the design equation is not simple due to the conditional probabilities. We propose a simpler approach, inspired by Flynn and Gray [4].

In this second algorithm, we define α_X as the composition of a primary quantizer Q_X and an index assignment (IA) function δ_X: $\delta_X : \mathcal{K}_X \to \mathcal{I}_X$, where $\mathcal{K}_X$ is the index set for Q_X. Similarly, we set $\alpha_Y = \delta_Y \circ Q_Y$. In the following, merging two quantization cells i and i' in $\mathcal{I}_X$ means creating a new IA function δ'_X identical to δ_X excepted that $\delta'_X(i) = \delta'_X(i') = \delta_X(i)$. We design the IA function by iteratively merging quantization cells until the current rate is equal to the target rate. We denote by $\Delta_{(i,i')}(.)$ (resp. $\Delta_{(j,j')}(.)$) the variation of the argument when i and i' (resp. j and j') are merged in δ_X (resp. δ_Y). We define the following *marginal returns*

$$\begin{aligned}\Lambda_X(i,i') &= -(\Delta_{(i,i')}(D_X) + \mu\Delta_{(i,i')}(D_Y))/\Delta_{(i,i')}(R_X) \\ \Lambda_Y(j,j') &= -(\Delta_{(j,j')}(D_X) + \mu\Delta_{(j,j')}(D_Y))/(\lambda\Delta_{(j,j')}(R_Y))\end{aligned}$$

where μ and λ are positive Lagrangian multipliers. We can find a good sequence of mergings by choosing at each step the pair of indices minimizing the corresponding marginal return. It amounts to choosing the merging that minimizes the distortion increase per bit. This approach is advantageous in that it can be implemented using a training set and it directly gives the whole rate-distortion curve. The drawback of the technique is the lack of optimality guarantee. As noted in [4], the re-computations of the rates and distortions after a merging can be made in time independent of the number of training samples.

III. Experiments

In these experiments, we defined X and Y as scalar Gaussian sources ($k = 1$) with unit variance and correlation factor $\rho = 0.9$. We first implemented the descent algorithm using numerical integration. We chose to train the encoder α_Y and the decoder β with – asymmetrically – α_X defined as a uniform quantizer with $R_X \approx 6.37$ bits. The encoder α_Y is initialized with an encoder identical to α_X. We then implemented the second algorithm (IA method). The encoder chosen for α_X is the same as above. For α_Y, we started with a uniform encoder with thinner intervals.

In the table below, we display the results for the descent algorithm, which confirm the advantage of entropy constraints (EC) over simple entropy limitations (non-EC).

R_Y	$D_X + D_Y$ (EC)	$D_X + D_Y$ (non-EC)	Difference
2.5	9.24×10^{-3}	11.09×10^{-3}	+0.79dB
3.5	2.34×10^{-3}	3.09×10^{-3}	+1.20dB
4.0	1.34×10^{-3}	1.59×10^{-3}	+0.75dB
5.0	5.17×10^{-4}	5.29×10^{-4}	+0.11dB

Then, similar results are shown for the IA method. In some cases, the method is really competitive to the more computationally intensive descent algorithm.

R_Y	$D_X + D_Y$ (EC)	$D_X + D_Y$ (non-EC)	Difference
2.5	9.46×10^{-3}	11.43×10^{-3}	+0.82dB
3.5	2.34×10^{-3}	3.36×10^{-3}	+1.56dB
4.0	1.52×10^{-3}	1.86×10^{-3}	+0.89dB
5.0	5.37×10^{-4}	6.49×10^{-4}	+0.82dB
6.0	2.91×10^{-4}	3.19×10^{-4}	+0.40dB

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, July 1973.

[2] Q. Zhao and M. Effros, "Optimal code design for lossless and near lossless source coding in multiple access networks," in *Proc. IEEE Data Compression Conf.*, 2001, pp. 263–272.

[3] M. Fleming, Q. Zhao, and M. Effros, "Network vector quantization," *IEEE Trans. Inform. Theory*, 2001, (submitted).

[4] T. J. Flynn and R. M. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, pp. 773–787, Nov. 1987.