Transform-based compression

Marco CAGNAZZO

1. Block quantization

If we quantize a uniform RV X (random variable) with a uniform quantizer, the resulting distortion is proportional to the RV variance and decreases exponentially with the rate (the "6 dB"-rule) :

$D = \sigma_X^2 2^{-2R}$ $D_i = \sigma_i^2 2^{-2R_i}$

More in general, for high-resolution uniform quantization of a generic RV, we have:

$D = K_X \sigma_X^2 2^{-2R}$

And for optimal quantization of a generic RV we have

$D = h_X \sigma_X^2 2^{-2R}$

Where h_X is called *shape factor* and only depends on the shape of the RV PDF. Eg., $h_X = 1$ for uniform RV and $h_X = \sqrt{3}\pi/2$for Gaussian RV.

Let us now consider a random vector $\boldsymbol{X} = [X_1\, X_2 \dots X_N]^T$. Let σ_i^2be the variance and h_i the shape factor of X_i. Let us suppose that each component of the input vector is quantized with an optimal quantizer. The resulting per-component distortion is:

$$D = \frac{1}{N}||\boldsymbol{X} - Q(\boldsymbol{X})||^2 = \frac{1}{N}\sum_{i=1}^{N} h_i \sigma_i^2 2^{-2R_i} = D(\boldsymbol{R})$$

Where $\boldsymbol{R} = [R_1\, R_2 \dots R_N]^T$is the rate allocation vector. The *optimal rate allocation* problem consists of finding the rate allocation vector minimizing the distortion subject to a total rate constraint:

$$\boldsymbol{R}^* = \arg\min_{\boldsymbol{R}} D(\boldsymbol{R})$$

Subject to $\sum_{i=1}^{N} R_i = R_T$

This problem is solved using Lagrange multipliers. The unconstrained cost function is:

$$J(\boldsymbol{R}) = \frac{1}{N}\sum_{i=1}^{N} h_i \sigma_i^2 2^{-2R_i} + \lambda\left(\sum_{i=1}^{N} R_i - R_T\right)$$

Computing the gradient of the cost function and putting it to zero, we can easily find:

$$R_i^* = \overline{R} + \frac{1}{2}\log_2 \frac{h_i \sigma_i^2}{h_{GM} \sigma_{GM}^2}$$

Where $\overline{R} = \frac{R_T}{N}$, $\sigma_{GM}^{2} = [\prod_{i=1}^{N} \sigma_i^{2}]^{1/N}$ and $h_{GM} = [\prod_{i=1}^{N} h_i]^{1/N}$. The subscript GM stands for *geometric mean*. This rate allocation is called Huang-Schultheiss formula.
Replacing the optimal rate value in the distortion expression, we find:

$$D^{*} = h_{GM}\sigma_{GM}^{2} 2^{-2\overline{R}}$$

That is, the distortion we would find quantizing an RV with variance σ_{GM}^{2}, shape factor h_{GM} with rate $\overline{R}$.
In the case of identically distributed (ID) RV, the Huang-Schulteiss allocation simply assigns the same rate to each component of the vector.

2. Transform coding

We hope to improve the rate-distortion curve of an input vector by performing the quantization in the transform domain with **transform coding**.
We consider orthogonal transforms, i.e., transforms that can be represented with orthogonal matrices (unitary matrices in the complex case):

$$\boldsymbol{Y} = A\boldsymbol{X}$$

The matrix A must be orthogonal, that is $A^{T} = A^{-1}$. DCT is an example of orthogonal transforms. This is equivalent to a rotation of the coordinate system and also implies $||\boldsymbol{Y}|| = ||\boldsymbol{X}||$
The transform coding paradigm consists of quantizing $\boldsymbol{Y}$and then computing the inverse transform:

$$\boldsymbol{X}_q = A^{-1}Q(\boldsymbol{Y})$$

The resulting per-component distortion is

$$D_X = \frac{1}{N}\left|\left|\boldsymbol{X} - \boldsymbol{X}_q\right|\right|^2 = \frac{1}{N}\left|\left|A^{-1}\boldsymbol{Y} - A^{-1}Q(\boldsymbol{Y})\right|\right|^2 = \frac{1}{N}(A^{-1}\boldsymbol{Y} - A^{-1}Q(\boldsymbol{Y}))^{T}(A^{-1}\boldsymbol{Y} - A^{-1}Q(\boldsymbol{Y})) = \frac{1}{N}(\boldsymbol{Y} - Q(\boldsymbol{Y})^{T}AA^{T}(\boldsymbol{Y} - Q(\boldsymbol{Y})) = D_Y$$

We have this nice result, that with orthogonal transform the distortion can be computed directly in the transform domain: $D_X = D_Y$
Therefore, if we quantize the components of the transformed vector using the Huang-Schilteiss allocation formula, the resulting distortion (computed on the transformed vector $\boldsymbol{Y}$) is:

$$D^{*}_{T,X} = D^{*}_{Y} = h_{GM,Y}\sigma_{GM,Y}^{2} 2^{-2\overline{R}}$$

That is, the transform-based distortion of X is obtained using the geometric mean of the **transformed** vector variances and shape factors. Therefore the transform coding approach may reduce distortion for the same rate **if the transformed vector has a geometric mean of variances and shapes smaller than those of the input vector**.
Now we introduce the transform coding gain (or *coding gain* for short) as the ratio between the distortion achieved with block quantization and the one achieved with transform coding:

$$G = \frac{D^{*}_{X}}{D^{*}_{T,X}}$$

That is also equal to the distortion reduction (if $G > 1$) when we use transform coding instead of block quantization. We easily find that:

$$G = \frac{h_{GM,X}\sigma_{GM,X}^2}{h_{GM,Y}\sigma_{GM,Y}^2}$$

We observe that the geometric mean of the variances (and of the shapes) of the transformed vector depends on the choice of the transform matrix.

Let us now consider the case where X is a zero-mean, Gaussian ID random vector. In this case, all the shape factors are identical (since Y is also a Gaussian random vector). Moreover,

$$\sigma_{GM,X}^2 = \sigma_X^2 = \sigma_{AM,X}^2$$

This is trivial: all the input vector components have the same variance, which is also equal to the geometric mean and arithmetic mean (AM). Now, we have that

$$\sigma_{AM,X}^2 = E(X_i^2) = \frac{1}{N}\sum_{i=1}^{N} E(X_i^2) = \frac{1}{N}E(\sum_{i=1}^{N} X_i^2) = \frac{1}{N}E(||\boldsymbol{X}||^2) =$$

$$\frac{1}{N}E(||\boldsymbol{Y}||^2) = \frac{1}{N}E(\sum_{i=1}^{N} Y_i^2) = \sigma_{AM,Y}^2$$

Finally, the coding gain becomes:

$$G = \frac{\sigma_{AM,Y}^2}{\sigma_{GM,Y}^2}$$

That is **the ratio of the arithmetic and geometric means of the transform coefficient variances.**

Therefore, given that all the orthogonal transforms share the same value of $\sigma_{AM,Y}^2$ (we just proved that $\sigma_{AM,Y}^2 = \sigma_{AM,X}^2$ for any orthogonal transform), **we have to look for transforms that reduce $\sigma_{GM,Y}^2$.**

The geometric mean of a set of positive real numbers with a given sum $N\sigma_{AM,X}^2$ is small if the numbers are as "different" as possible. Ie, the optimal transform should create a few transform coefficients with a very large variance and many coefficients with a very small variance. This is also called "energy concentration".

It can be shown that DCT is asymptotically optimal for Gaussian random vectors in terms of energy concentration.

Exercise

Theoretical part: prove that the Huang-Schultheiss formula is the solution of the constrained rate allocation problem.