NETWORKS OF THE POOLING TYPE AND OPTIMAL QUANTIZATION

Pierre-Olivier Amblard, Steeve Zozor *

Laboratoire des Images et des Signaux
CNRS UMR 5083, Grenoble
LIS-BP46, 38402 Saint-Martin d'Hères, France

Olivier J. J. Michel †

Laboratoire Universitaire d'Astrophysique de Nice
Univ. Nice-Sophia-Antipolis, UMR CNRS 6525
Parc Valrose, 06108 Nice cedex 2, France

ABSTRACT

We study the link between networks of the pooling type and the problem of quantification. Pooling networks consist in parallel processors that are summed after having processed the same information. If the processors are simple threshold model of noisy neurons, the behavior of the network has the behavior of a quantizer. Using the compander approach to quantizers as well as the notion of density of levels, we study these networks and show that they are asymptotically equivalent to quantizers. Furthermore, we show how these devices can be infomax processors.

1. INTRODUCTION

Pooling networks of neurons are a special type of neural networks that process N times the information and merge the processing in some way (usually an average). This architecture, depicted in figure (1) occurs in many parts of the sensory pathway of the human (e.g. in the auditory pathway, in the visual pathway [1, 2, 3, 4],...). Each neuron can be modeled using sophisticated Hodgkin-Huxley equations, but to be able to study the flow of information, we adopt a very simplified model: a neuron is a noisy nonlinear threshold model. This simplification allows to incorporate the fluctuations of the neural system as well as one of the fundamental feature of neurons, the threshold. Indeed, a neuron sends information when one of its state variables, the membrane voltage, exceeds some threshold. Further, this particular architecture also exists in some engineering fields: dimus arrays in sonar [5, 6], flash analog-to-digital converters [7] are some examples.

Pooling networks have been studied in the physics literature [8, 9]. They possess the strange property of noise-enhanced processing, meaning that the performance of the processing of information has a non monotonic behavior as a function of the fluctuations intensity. Therefore, there exists an optimal power of the fluctuations for which the processing is optimal. Furthermore, as mentioned in [8], there is a close relationship between pooling networks and quantization. Indeed, the conjunction of the additive noise term n_i and the threshold θ can be viewed as a random quantization on 1 bit. If N is the size of the networks, summing leads to a random quantization over N bits. Quantizers are nonlinear settings that discretize the amplitude of an input. In most practical situations, a quantizer can be decomposed as the cascade of a compressor nonlinear device that adapt the dynamic of the input, followed by a uniform quantization and an expansion nonlinearity that brings back the amplitude in the dynamic range of the input. This decomposition is called the compander model and is illustrated in figure (1).

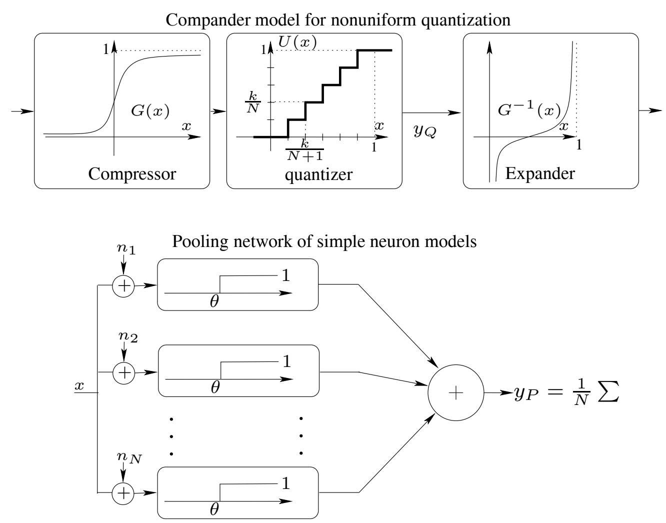

Fig. 1. The compander representation of a scalar quantizer (top). Pooling networks of noisy threshold devices are simple models for pooling networks of real neurons (bottom).

The aim of the paper is to go further in the analogy between pooling networks and quantization, using the compander model for quantizers. We show the asymptotic equivalence in law of the devices, and derive asymptotic input-output information theoretic based relationships which allow to stress the importance of the noise in the network and of the compressor in the quantizer.

In the following section, we recall some facts on quan-

*bidou.amblard,steeve.zozor@lis.inpg.fr
†olivier.michel@unice.fr

tizers and pooling networks, facts which help us in section 3 to show the asymptotic equivalence of the two devices. The equivalence is studied in terms of the probabilistic structure of the output as well as information theoretic arguments.

2. SOME FACTS ON QUANTIZERS AND POOLING NETWORKS

We recall here some known facts on quantizers and pooling networks. This allows the setting of some notations and introduce some concepts for unfamiliar readers.

2.1. Quantizers

Scalar quantizers are defined by a set of thresholds $t_i, i = 0, \ldots, N$ and a set of levels $q_i, i = 1, \ldots, N$ such that the quantization of a variable x is $y = Q(x) = q_i \mathbf{1}_{(t_{i-1}, t_i]}(x)$ where $\mathbf{1}_I(x)$ stands for the characteristic function over interval I. By convention, $t_0 = -\infty$ and $t_N = +\infty$. Furthermore, we will consider here regular quantizers for which $q_i \in (t_{i-1}, t_i]$. For this type of quantizers, it is easy to show that Q can be decomposed as $Q = G^{-1} \circ U \circ G$ where U is a uniform quantizer ($t_i - t_{i-1} = \Delta, q_1 = t_1 - \Delta/2, q_i - q_{i-1} = \Delta$), G a bijective function which can be chosen smooth enough, and which compresses the dynamics of the input, and $\circ$ is the composition of functions. The inverse of G is called an expander, and the full decomposition is known as the *compander* approach to quantization [10].

An important aspect of quantizers is the asymptotic theory, also know as fine or high-rate quantization [10, 11, 12]. In this approach, the number of levels grows to infinity, and the levels are described by a density instead of their precise location. The density of levels is assumed to exist and is defined as $\lim_{N \to +\infty} \{\# q \in [q, q + dq)\}/N = \lambda(q)$ (# for number of). There is a close relationship between density λ and the compressor function G. Indeed, G sends the interval $(t_{i-1}, t_i]$ of length Δ_i onto an interval of fixed length Δ, and therefore $G(t_i) - G(t_{i-1}) = \Delta = T/N$ where T is the quantizer dynamical range. But in an interval of length Δ_q, there are about $\lambda(q) N \Delta_q$ levels and therefore in this interval, the length between levels is approximately $\Delta_i = 1/N\lambda(q)$. Therefore, as N grows we get $G'(q) = \Delta/\Delta_i = T\lambda(q)$: the compressor is the cumulative density function of density λ (if T=1).

2.2. Pooling networks

Pooling networks have been extensively studied by N. Stocks. The probabilistic structure of the output is quite simple to obtain. Indeed, the N variables n_i are assumed independent and identically distributed. The output y_P is then conditionally binomially distributed $P(y_P = k/N|x) = \binom{N}{k} P_n(x, \theta)^k (1 - P_n(x, \theta))^{N-k}$ ($P(A)$ stands for probability of event A). Therefore $P(y_P = k/N) = E_x[P(y_P = k/N|x)]$ where $E_x[.]$ stands for the mathematical expectation over x, $\binom{N}{k}$ is the binomial coefficient, and where $P_n(x, \theta) = P(n + x \geq \theta)$ is the probability that the noisy input is over the threshold. Equipped with these results, it is easy to evaluate the Shannon mutual information between the input and the output $I(x, y_P)$ which can be given by the difference between the entropy $H(y_P)$ of the output and the conditional entropy $H(y_P|x)$. But, no close form solution exists for the entropy nor the mutual information, and theoretical optimization with respect to the input probability structure (*capacity* of the network) or the parameters of the neurons (*infomax* processor) is not possible to perform. However, if the probability density functions p_x of x and p_n of n are even, a symmetry argument shows that θ should be chosen equal to 0. This choice is adopted in the following.

3. POOLING NETWORKS AND QUANTIZERS ARE ASYMPTOTICALLY EQUIVALENT

To compare pooling networks and quantizers, we adopt the compander approach and only consider the first two stages $U \circ G$. We assume that the output of the uniform quantizer takes its values among the $N + 1$ values $q_i = i/N, i = 0, \ldots, N$, q_i being chosen if the input lies in $(t_i = i/(N + 1), t_{i+1} = (i+1)/(N+1)$. Note that this quantizer is not strictly uniform because the quantization step $1/N$ is different from the time step $1/(N + 1)$. However, it is asymptotically uniform, and the compander decomposition is still valid, even at finite N.

3.1. Probabilistic aspects

Let y_Q denotes the output of the uniform quantizer, or $y_Q = U \circ G(x)$. The probability law of the random variable is then

$$P\left(y_Q = \frac{k}{N}\right) = P\left[G(x) \in \left(\frac{k}{N+1}, \frac{k+1}{N+1}\right]\right]$$
$$= F_x \circ G^{-1}\left(\frac{k+1}{N+1}\right) - F_x \circ G^{-1}\left(\frac{k}{N+1}\right) \quad (1)$$

where $F_x(u)$ is the cumulative density function (c.d.f.) of x.

Let y_P the output of the pooling network. We have seen that the probability law of y_P is

$$P\left(y_P = \frac{k}{N}\right) = E_x\left[\left(\binom{N}{k} F_n(x)^k (1 - F_n(x))^{N-k}\right]\right. \quad (2)$$

where $F_n(u)$ is the c.d.f. of n, and where the evenness of the densities is assumed.

Let us now focus on the asymptotic regime for which N, the number of levels and the number of neurons, goes to infinity. For the quantizer, this amounts to consider the high-rate regime, and for the network, this will give the behavior of large networks as encountered *e.g.* in the brain.

The probability law of the output of the network is given by eq. (2), or

$$P\left(y_P = \frac{k}{N}\right) = \int_{\mathbb{R}} \binom{N}{k} F_n(u)^k (1 - F_n(u))^{N-k} p_x(u) du$$

If we perform the change of variable $v = F_n(u)$, we get

$$P\left(y_P = \frac{k}{N}\right) = \int_0^1 \binom{N}{k} u^k (1-u)^{N-k} g(u) du$$
$$\text{where } g(u) = \frac{p_x \circ F_n^{-1}(u)}{p_n \circ F_n^{-1}(u)}$$

Let $B(x,y) = \int_0^1 u^{x-1}(1-u)^{y-1} du = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ be the Beta function [13] ($\Gamma(x)$ is the Gamma function). Since for an integer $n! = \Gamma(n+1)$, we get

$$P\left(y_P = \frac{k}{N}\right) = \frac{1}{N+1}\int_0^1 \frac{u^k(1-u)^{N-k}}{B(k+1, N-k+1)} g(u) du$$

Function $\delta_{N,k}(u) = \frac{u^k(1-u)^{N-k}}{B(k+1,N-k+1)}$ is a probability density function defined over $[0,1]$, and can be shown to behave as a delta function at the point k/N as N grows. Indeed, it is monomodal and maximal at k/N; its value at the maximum goes to infinity with N, and as a probability density function, it sums to one. Therefore, we have the following result

$$P\left(y_P = \frac{k}{N}\right) \sim_N \frac{1}{N+1} g\Big(\frac{k}{N}\Big) \quad (3)$$

where $a_n \sim_n b_n \Leftrightarrow \lim_{n\to+\infty} a_n/b_n = 1$. Note that a more rigorous derivation of this result can be obtained using the Laplace method to approximate integrals.

The probability law of the output is given by eq. (1). Using the mean value theorem, there exists $c_N^k \in (k/(N+1), (k+1)/(N+1)$ such that

$$P\left(y_Q = \frac{k}{N}\right) = \frac{1}{N+1}(F_x \circ G^{-1})'(c_N^k)$$

But $(F_x \circ G^{-1})' = (p_x \circ G^{-1})/(G' \circ G^{-1})$, and therefore, if we choose $G = F_n$ we obtain

$$P\left(y_Q = \frac{k}{N}\right) = \frac{1}{N+1} g(c_N^k)$$

It is easy to see that $c_N^k \sim_N k/N$ (both belong to interval $(k/(N+1), (k+1)/(N+1))$ the length of which goes to zero when N goes to infinity) and therefore we conclude that $P\left(y_Q = \frac{k}{N}\right) \sim_N P\left(y_P = \frac{k}{N}\right)$ if of course $G = F_n$.

3.2. Information theoretic aspects

We now focus on the transmission of information through the devices.

It has been shown [8] and it is easy to verify that $P(y_P = k/N) = 1/(N+1)$ whenever $F_n(u) = F_x(u)$. In this case, the network equalizes the input, and the output is of maximum entropy $H(y_P) = \log(N+1)$. Therefore, *if the threshold distribution equals the distribution of the input and if the compressor is the c.d.f. of the input, the quantizer and the pooling network are equal in law.* Furthermore, as we saw previously, this situation corresponds for the quantizer to the infomax situation, whatever N: This is not true for the network since $H(y_P|x) \neq 0$. This is illustrated in figure (2.B) where we plot the mutual information in the case of zero mean Gaussian laws for x and n, as a function of parameter $s = \sigma_n/\sigma_x$, where σ_z^2 is the variance of variable z. As clearly seen in the figure, the maximum of the mutual information for the network does not occur at $s = 1$, and hence not for $F_n(u) = F_x(u)$. However, it seems that as N grows, Arg $\max_s I(x, y_P)$ grows. The question is then, does Arg $\max_s I(x, y_P) = 1$ when $N \to +\infty$?

The mutual information between the input and the output of the quantizer reads $I(x, y_Q) = H(y_Q)$ since the device is deterministic. Therefore, the *infomax* quantizer is the quantizer which maximizes the entropy of its output. We have seen that $P(y_Q = k/N) = g(c_N^k)/(N+1)$ and therefore the entropy is maximal if and only if $g(c_N^k) = 1, \forall k = 0, \ldots, N$. Remember that $g = p_x \circ F_n^{-1}/p_n \circ F_n^{-1}$ is the compressor is chosen equal to F_n. Therefore, for finite N a sufficient condition to ensure y_Q to be of maximal entropy is $p_x = p_n$ or $G = F_x$. This condition is necessary if N goes to infinity.

The entropy of the output and the mutual information in the case of the network are more tedious to evaluate. Since the p.d.f. involved are assumed even, it easy to show

$$I(x, y_P) = 2N\int_0^1 u g(u)\log(u) du + H(y_P) + \sum_{k=0}^{N} P\Big(y_P = \frac{k}{N}\Big)\log\binom{N}{k},$$

a formula which can be hardly pushed any further exactly. However, using the equivalence (3) for the probability law at the output of the network, we are able to evaluate an approximation for the discrete sums in the previous equation. The calculation relies on the Riemann approximation of integrals. We obtain for the entropy the following asymptotic approximation

$$H(y_P) = \log(N+1) - D_{KL}(p_x||p_n) \quad (4)$$

where $D_{KL}(p_x||p_n) = E_x[\log p_x/p_n]$ stands for the Kullback divergence between densities p_x and p_n. This last result shows how far the output of the network is from its maximum entropy position. It also shows that the output has maximum entropy if and only if the density of the noise is equal to the density of the input, a result previously recalled in the paper. The entropy is depicted in fig. (2.A) for several values of N in the Gaussian case. The approximation is very good for $s \geq 1$ but becomes poorer and poorer as s goes to zero. This is due to the fact that for $s < 1$, function g diverges at 0 and 1: for $s < 1$ the error in the approximation crucially depends on s, and N must be very large in order to compensate the divergence of g in the numerical calculation.

For the mutual information, we obtain the approximation

$$I(x, y_P) \quad = \log \frac{(N+1)!N^N}{e^N (N!)^2} - D_{KL}(p_x || p_n). \qquad (5)$$

This shows also that the mutual information will be maximized for $p_x = p_n$ but **only asymptotically**, and therefore for parameter s equals to 1 in the Gaussian case. Furthermore, the control of the error is even more difficult in the case of the mutual information since it appears to be the difference of many approximated terms. For the Gaussian case, the approximation validity crucially depends on s, and the lower s the larger N for the approximation to hold. We have to work further on the control of the error of the approximation using both variable N and s. Finally, if the approximation allows to say that the network is infomax at $s = 1$ when $N \to +\infty$, it however too crude to allow the computation of the curve $s_N = f(N)$ of the maximizing parameter s. We need to develop a more precise approximation.

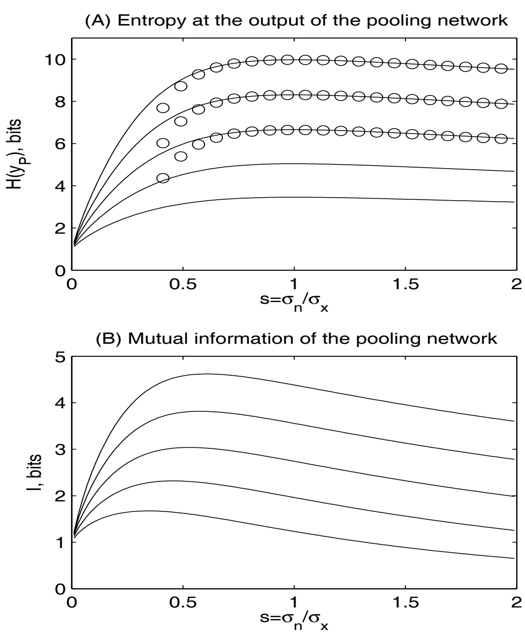

Fig. 2. Plotted against the leading parameter $s = \sigma_n/\sigma_x$ for different values of the size N of the network (up to bottom: N=10, 32, 100, 316, 1000). (A) Entropy of the output of the pooling network. We have superposed the plot (circles) of the approximation (4) for N=100, 316, 1000. (B) Mutual information of the pooling network.

4. CONCLUDING REMARKS

We have shown here the full asymptotic equivalence between pooling networks of noisy neurons with quantizers. Precisely, when the compressor function of the quantizer is chosen to be the c.d.f. of the density of the noise in the network, the output of the quantizer and of the network are equal in law asymptotically. This allows to see the network as a random sampler of the quantizer, and it also open perspectives for the study of particular properties of the network

We have developed asymptotic approximation for the entropy and the mutual information which allows to formulate some already know conclusion. In order to be more useful, these crude approximations needs to be tightened.

Our approach will be generalized to the case of different thresholds. This case is particularly interesting since the optimization of the thresholds creates a structure of bifurcation among the thresholds as a function of parameter s [14]. This particular structure is not fully understood and our point of view may lead to improvement in these questions.

5. REFERENCES

[1] F. RIEKE et. al., *Spikes?Exploring the neural code*, The MIT Press, 1996.

[2] P. DAYAN and L. F. ABOTT, *Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems*, The MIT Press, 2001.

[3] W. GERSTNER and W. KISTLER, *Spiking Neuron Models, single neurons, populations, plasticity*, Cambridge University Press, 2002.

[4] N. G. STOCKS, D. ALLINGHAM, and R. P. MORSE, "The application of suprathreshold stochastic resonance to cochlear implant coding," *Fluct. Noise Letters*, vol. 2, pp. L169–L181, 2002.

[5] V. C. ANDERSON, "Digital array phasing," *J. Acoust. Soc. Am.*, vol. 32, pp. 867, 1960.

[6] P. RUDNICK, "Small signal detection in the dimus array," *J. Acoust. Soc. Am.*, vol. 32, pp. 971, 1960.

[7] V. K. MADISETTI and D. B. WILLIAMS, *The digital signal processing handbook*, CRC Press, Boca Raton, Fl., 1998.

[8] N. G. STOCKS, "Information transmission in parallel threshold arrays: suprathreshold stochastic resonance," *Phys. Rev. E*, vol. 63, no. 041114, 2001.

[9] N. G. STOCKS and R. MANELLA, "Generic noise-enhanced coding in neuronal arrays," *Phys. Rev. E*, vol. 64, no. 030902(R), 2001.

[10] A. GERSHO and R.M. GRAY, *Vector quantization and signal compression*, Kluwer Academic Publishers, 1991.

[11] R. M. GRAY and D. L. NEUHOFF, "Quantization," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[12] S. GRAF and H. LUSCHGY, *Foundations of quantization for probability distributions*, LNM 1730, Springer-Verlag, 2000.

[13] I.S. GRADSHTHEYN and I.M RYSHIK, *Tables of integrals, series and products, 4th ed.*, Academic Press, 1980.

[14] M. D. McDONNELL, N. G. STOCKS, C. E. M. PEARCE, and D. ABBOTT, "Point singularities and suprathreshold stochastic resonance in optimal coding," ArXiv/cond-mat/0409528 2004.