STABILITY OF MULTILAYERED NEURAL NETWORKS

M. Miller and E. N. Miranda
Centro Atomico Bariloche and Instituto Balseiro.
8.400, S. C. de Bariloche, Argentina.

Abstract: Stability of multilayered neural networks against synaptic changes has been studied numerically. We have found that the average maximum change goes to zero as the number N of input neurons is much greater than one. If a fixed fraction of output errors is allowed, then the synapses may be changed within some limits even for large N.

The connectionist approach to Artificial Inteligence problems has been very fruitful [1]. In this context multilayered neural networks have been intensively studied recently. However, little is known about general properties of these models, although there are some results about learning times in these networks [2]. It is important to know how the typical number of steps needed to learn a task varies with the system parameters. Also, the learning process has been associated with a properly defined entropy [3]. But there are other problems which have not been studied yet. For example, little is known about the metastable states in configuration space. In this letter, the stability of the networks against changes of synapses is studied both numerically and analyticaly. How much can be changed the strength of a synapses without altering the output of the network? In order to answer this question, we performed a numerical study and we explained the results with a simple probabilistic model.

We find that the average maximum change goes to zero as the number N of imput neurons increases. This result is intuitively correct since the system is forced to give the right answer for the whole set of 2^N different examples. This constraint could be relaxed: a finite fraction α of the total number of examples, $\alpha 2^N$, is allowed to be retrieved incorrectly. In such a case, the average maximum modification reaches a limit value.

A specific architecture with N input neurons, L hidden units and one output neuron was studied. The first and the second layers have also a neuron S_0 which is kept always active and acts as a threshold for the neurons in the next layer. The synapses J are chosen with a uniform probability

distribution in the interval [-K,K]. A neuron is connected with all neurons of the following layer.

The hidden neurons are continuous variables in the range [-1,1]. J_{ij} is the synapse between neuron j of the input layer and neuron i of the hidden layer; h_i is the internal field on the ith neuron of the hidden layer and H_i is its threshold. Their response function is:

$$S_i(t+1) = tgh\left[h_i + H_i\right] = tgh\left[\sum_{j=1}^{N} J_{ij} S_j(t) + H_i\right]$$

Then, K was set at the value K=10 to keep the hidden neurons almost always saturated at ±1. Since the distribution is symmetric, its average value is zero and its variance is σ_o =5.77. The thresholds are chosen with the same distribution.

The output neuron response function is:

$$S(t+1) = sign\left[h + H\right] = sign\left[\sum_{j=1}^{L} J_j S_j(t) + H\right]$$

J_j is the synapse between neuron j of the hidden layer and the output neuron; h is the internal field on the output neuron and H is its threshold.

No learning mechanism was considered. The networks are chosen randomly and the right answer for each input is that given spontaneously by the system.

Two different ensembles were studied. In the ensemble A, the networks are chosen randomly; then it is checked explicitly that all hidden neurons are used. This means that there should be at least one error in the output when each hidden neuron is put equal to zero. If a network has

unused hidden neurons, then it is not accepted. If there were unused neurons in the network, then there would be a fictitious stability since some synapses could support large changes without producing errors.

The ensemble B is a completely random one and the networks are always accepted. The stability of these networks can be understood in terms of simple probability considerations due to their random nature. Ensembles A and B are equivalent under certain conditions as it will be shown below.

Once a network has been chosen and tested, the following numerical experiment was done. Each internal synapse -the J_{ij} - is studied. Its value is changed in a certain number n of steps Δ until an output error occurs. In this way, the half width ($n\Delta$) associated to the synapse J_{ij} is measured. This procedure is repeated with steps $-\Delta$ and the number n' of allowed changes without errors is determined. So the width A associated to the synapse J_{ij} is: $(n+n')|\Delta|$. The whole procedure is repeated for the other synapses. In this way, a synapse width probability distribution $P_i(A)$ is obtained. The external synapses $-J_j-$ can be studied in the same way obtaining a probability distribution $P_e(A)$. The experiment was repeated over many different networks -from several hundreds to few thousands, depending on N and L-. The average A is used to characterize the measured distribution:

$$\overline{A_{i,e}} = \int P_{i,e}(A)\, A\, dA$$

In figure 1a and 1b, $\overline{A}$ and $\overline{A}$ are shown for different values of N in terms of the number L of hidden neurons. It

is clear for networks of the ensemble A that the internal synapses width increases with L -see figure 1a-. External synapses show the opposite behaviour in L: their width decrease with increasing L. In both figures, the width value goes to zero as N increases. Then, the network as a whole becomes unstable against changes in its synaptic strength as N increases.

It is very hard to understand what happens with networks of ensemble A because of the way they are chosen. So, we considered the same numerical experiment for networks of ensemble B. In this case, simple probabilistic considerations may be done and a theoretical model may be built which explains well the numerical data. It will be shown below that, under some conditions, both ensembles are almost equivalent.

The model is based on two hypothesis: 1) the internal field on a neuron is a gaussian distributed random variable; 2) a sign change in the internal field implies a change in the output.

A numerical test of these hypothesis shows that (1) is correct if the number of terms contributing to the internal field is greater than 5. Obviously (2) is right for the output neuron; but we have found it is wrong for hidden neurons. Then, one expects the model could work well for the external synapses.

The internal field on the output neuron is

$$h = \sum_{k=1}^{L} J_k S_k + \theta$$

It is assumed that the S_K takes the values ± 1 with equal probability. Then h follows a gaussian distribution

centered around the threshold H; its variance is:

$$\sigma = \sqrt{L}\,\sigma_0$$

Then, the internal field distribution on the output neuron is:

$$P(h) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ - \frac{(h-H)^2}{2\sigma^2} \right\}$$

The width distribution associated with the external synapses is determined as follows. Let $r_H(i)$ be the probability that h lies in the interval $(i-1)\Delta < h < i\Delta$ for a certain threshold H. N_{ex} is the number of different configurations of hidden neurons, then there will be N_{ex} values of h. The half width of the synapses is determined by the value of h closest to zero. The probability $p_H(j)$ that this h lies in the interval j is:

$$p_H(j) = \sum_{i=1}^{j-1} \left[1 - r_H(i)\right]^{Nex} \cdot \left[1 - \left(1 - r_H(j)\right)^{Nex}\right]$$

This is the probability that the synapses half width is $j\Delta$. The probability that the total width of the synapses is $k\Delta$, may be computed by:

$$\wp_H(k) = \sum_{j=0}^{k} p_H(j) \cdot p_H(k-j)$$

Then the average width is:

$$\overline{A}_{e,H} = \sum_{k} \wp_H(k) \cdot k\Delta$$

All of this computations are valid for synapses that reach a neuron with a certain threshold value H. The measured average width may be obtained by averaging over all possible values of H:

$$\overline{A}_e = \int_{-K}^{K} \overline{A}_{e,H}\, dH \Big/ \int_{-K}^{K} dH$$

N_{ex}'s value is bounded by 2^L since this is the maximun number of configurations of L neurons. If L > N then $N_{ex} \cong 2^N$ since each different input would produce a different internal configuration with hight probability. The case L < N is not interesting.

The average external width as a function of L for N < 5 was computed. The theoretical model fits well the data from ensemble B. Those of ensemble A have a very different behaviour. The difference, however, is due to finite size effects. For a larger number of input neurons, both ensembles behave in a similar way. This fact can be infered from figure 2 where the average external width is plotted in terms of L for N=7 for the ensemble A (+) and B (o). The full line is the model prediction.

It is easy to understand why both ensembles give similar results for large N. If a network belongs to ensemble B but not to ensemble A, this means there is at least one irrelevant hidden neuron. This means that if it is set to zero, no output mistake occurs for any of the 2^N different inputs. The average number of irrelevant hidden neurons goes as $L/2^N$. So if N >> lnL/ln2, then the number of networks with irrelevant neurons is negligible and both ensembles give the same result. The average fraction of networks with irrelevant neurons as a function of N for a fixed value of L can be seen in figure 3. It is clear that this number goes to zero with increasing N. So both ensembles are equivalent if the above condition is fulfilled.

From the data shown in figure 1, it is obvious that the network stability goes quickly to zero as N increases because a very small change in the external synapses cause a

change in the output. This result is intuitively correct since the network is asked to give the right answer for all the 2^N input examples. This constraint may be relaxed and a certain fraction α of mistakes may be accepted. In this case, one expects there should be an stability area in the synapse space.

A numerical experiment was carried out to test the above considerations. The technical details are as in the previous one. The only difference is that a synapse value is changed until there are $\alpha\, 2^N$ output mistakes. The results are shown in figure 4. The average external synapse width is plotted against N for a fixed value of L (L=10), when no mistake is accepted (o) and when 0.1×2^N mistakes are allowed (+). It is clear that in the last case, the synapse width reaches a limit value different from zero.

In summary, the stability of a simple layered neural network architecture has been studied numerically. The results can be understood in the limit of $L > N > \ln L/\ln 2$ in terms of simple probabilistic considerations. If no mistake is allowed, then the stability goes to zero for $N \to \infty$. If a fixed fraction of output mistakes is allowed, the synapses may be changed within some limits.

The authors thank N. Parga for discussions and a careful reading of the manuscript and S. Solla for discussions at the beginning of this work.

REFERENCES:

[1] D. E. Rummelhart and J. L. McLelland; "Parallel Distributed Processing", MIT Press, 1986, Cambridge. (MA).

[2] K. A. Benedict; 'Learning in multilayer perceptrons' J. Phys. A21 (1988) 2643

[3] P. Carnevali and S. Patarnello; 'Exhaustive Thermodinamical Analysis of Boolean Learning Networks'. Europhys. Lett. 4(1987) 1199

FIGURE CAPTIONS

Figure 1: Average maximun change of the internal (a) and external (b) synapses as a function of the number L of hidden neurons for different values N of input neurons. The networks as a whole become unstable against synaptic changes for large N due to their external ones.

Figure 2: External synapses average maximum change as a function of L (N=7) for networks of ensemble A (+) and B (o). The full line corresponds to the theoretical model which fits very well the data of ensemble B for any value of N. If N>5, the model also fits the data of ensemble A.

Figure 3: Fraction of networks of ensemble B which do not belong to ensemble A as a function of N for L=10. It is clear why the theoretical model also works well for ensemble A.

Figure 4: External synapses average maximum change as a function of N for L=10. If no ouput mistakes are allowed (o), the stability goes to zero with increasing N. If a fixed rate of mistakes (0.1×2^{N}) is permitted, then the stability reaches a limit value (+).

Fig 1 a

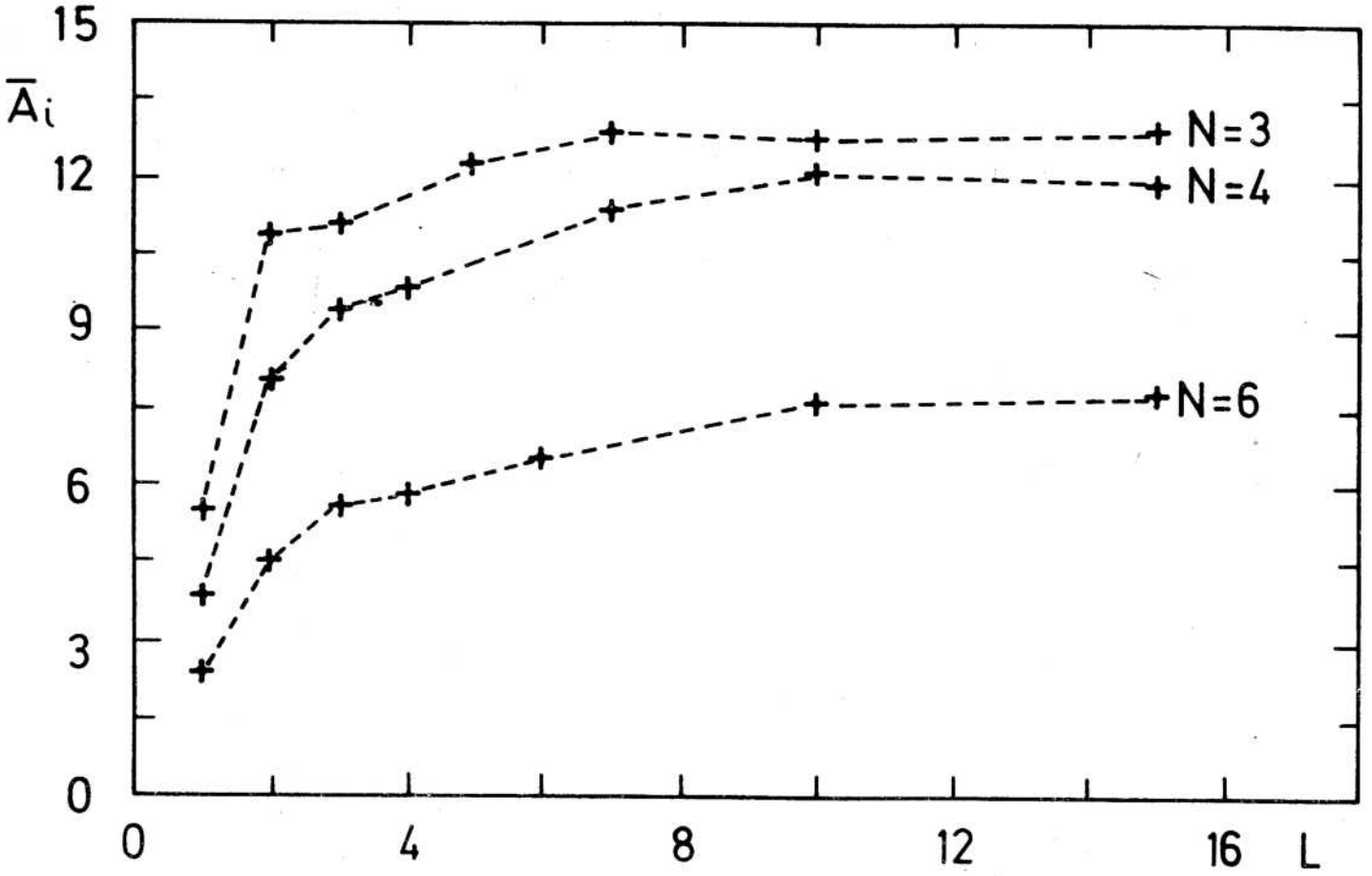


Fig. 1b

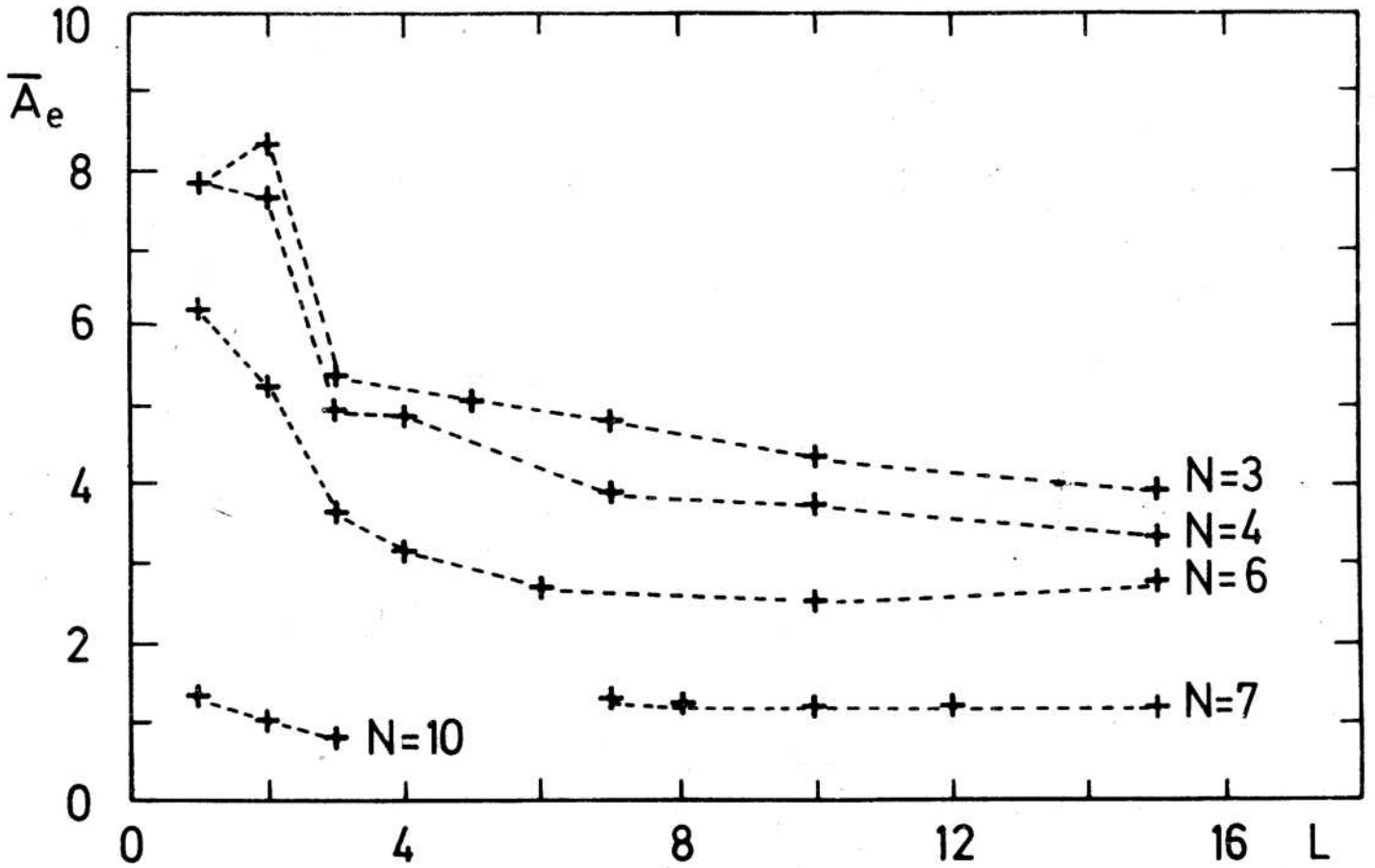


Fig. 2

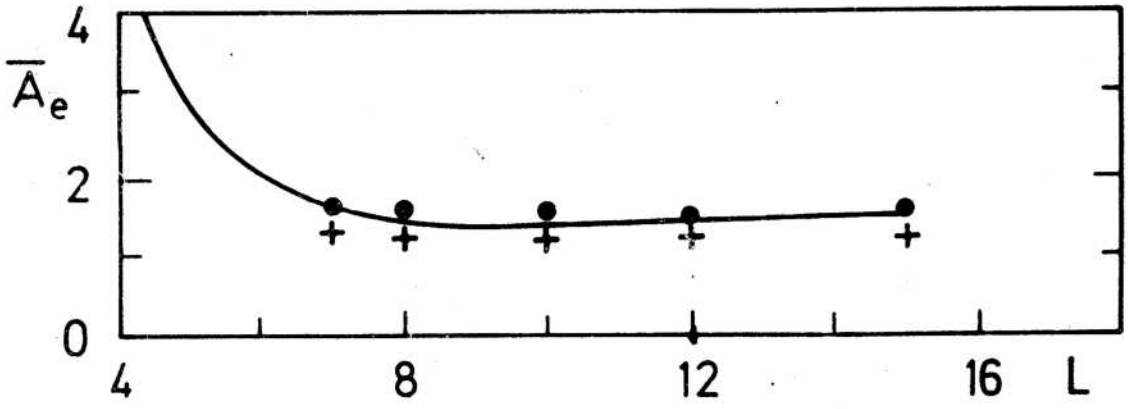


Fig. 3

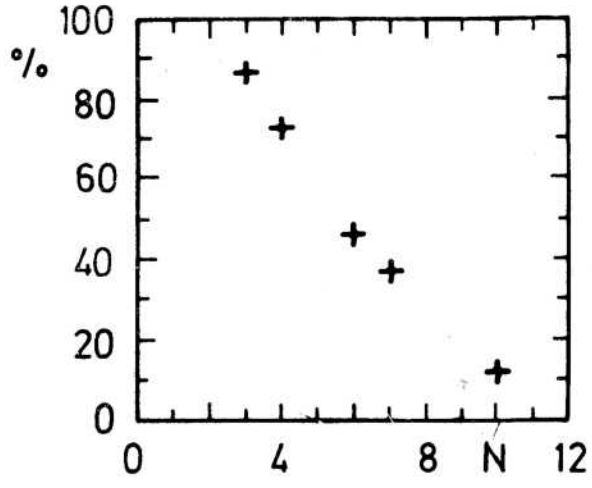
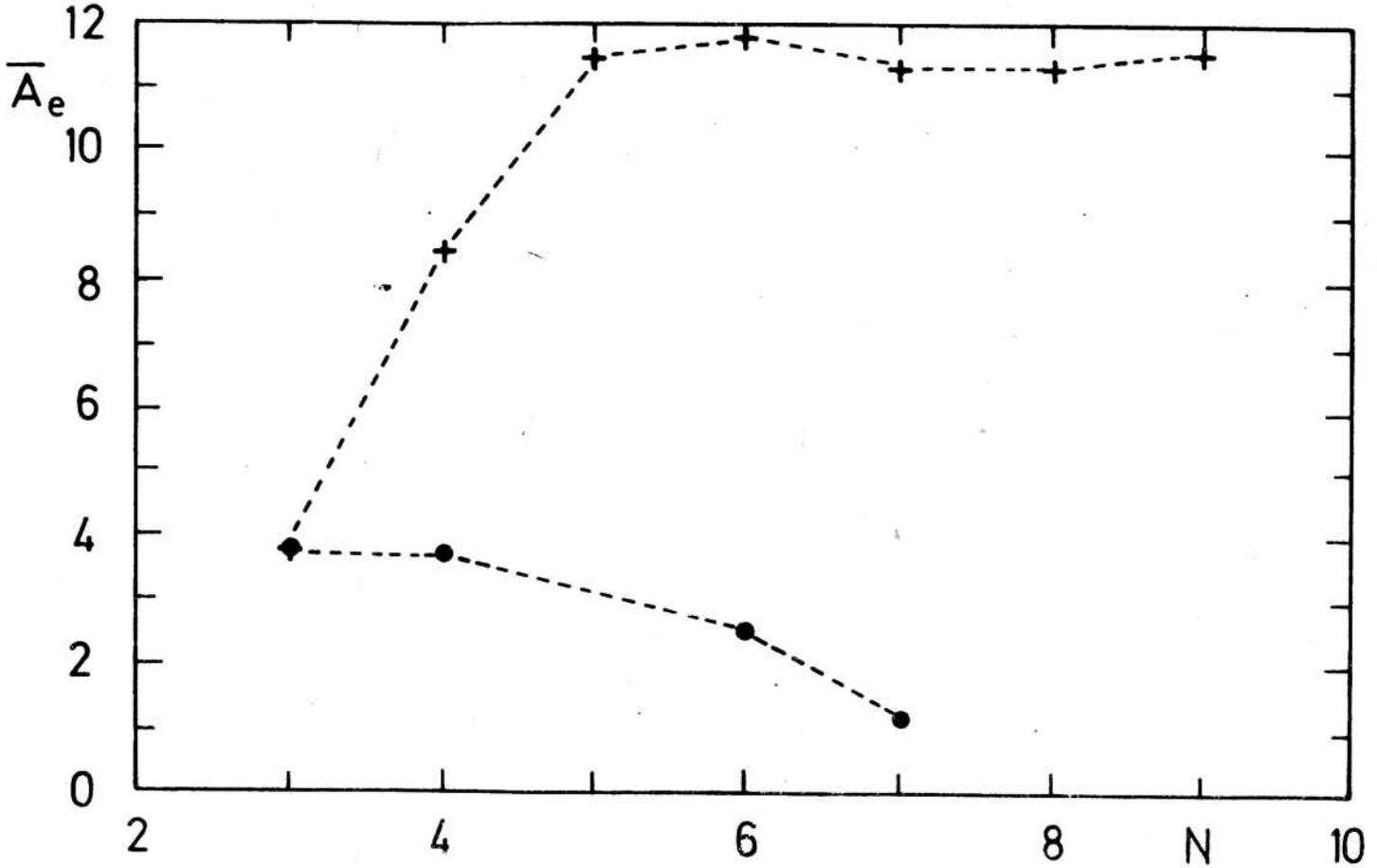


Fig. 4



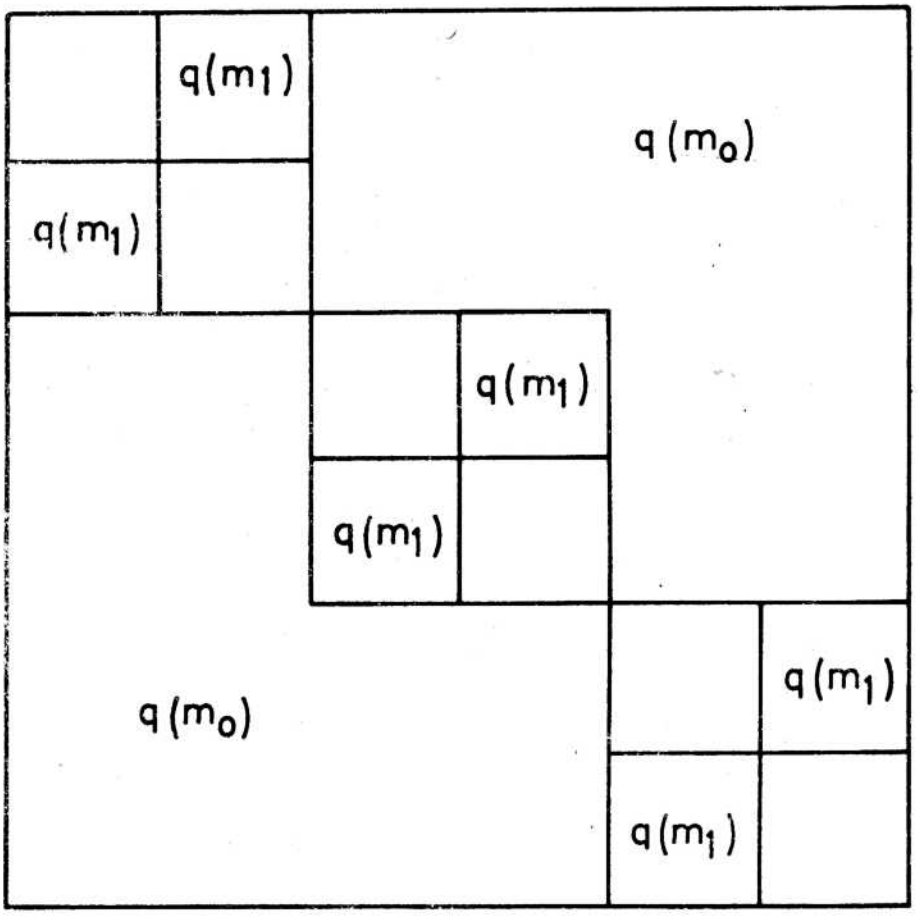


Fig 4