Distributed Optimization with the ALOPEX Algorithms

I. Marsic and E. Micheli-Tzanakou

Department of Biomedical Engineering, Rutgers University
Piscataway, NJ 08855-0909

Abstract. A method of distributed optimization using the ALOPEX algorithm and a paradigm of temperature spreading is described. It shows several desirable properties: the number of iterations is significantly smaller than with the global optimization and independent on the data set size; a stopping condition is introduced; the biological plausibility for neural network implementations is also maintained. In particular, applications to the problem of pattern recognition are presented.

1 Introduction

All current implementations of stochastic optimization techniques need to compute one global cost function in every iteration [5,8]. The result of this is that it is not easy to implement parallel versions of these algorithms. The number of iterations to get convergence to the optimal value depends on the data set size, and grows significantly with the dimension of the data set size. In the literature, there are some attempts of global optimization techniques without explicit calculation of a global cost function [3,4,7].

Our goals consist of: (a) to reduce the number of iterations needed to get convergence, and (b) to obtain global optimization using local processes.

2 Algorithm Implementation

The neural network algorithm consists of an input array where the data are presented and two layers of "neurons" (Fig.1). The input array is a hexagonal lattice, containing the image which should converge to one of the stored patterns. The neuron layers are:

1. "Feature Analyzers" (F_j) with hexagonal receptive fields connected to the input array. If s templates are memorized, then each feature cluster has s F_j's, each belonging to a particular stored template. The receptive fields of the feature clusters are overlapping to allow a spreading of "temperature". The temperature is defined as the inverse of the response strength.
2. "Response Integrators" (A_k) compute local cost functions. In this layer, there are also lateral connections between neurons. These connections are mutually inhibitory. This is necessary in order to ensure temperature transfers from warmer to colder regions.

The update of a particular variable in the t-th iteration is computed as in the original ALOPEX algorithms [1,2,8]:

$$x_i(t) = x_i(t-1) + \Delta x_i(t) = x_i(t-1) + \gamma \Delta x_i(t-1) \Delta R_k(t-1) + r_i(t), \quad \ldots (1)$$

where $\Delta x_i(t-1) = x_i(t-1) - x_i(t-2)$, $\Delta R_k(t-1) = R_k(t-1) - R_k(t-2)$, and $r_i(t)$ is random number. The cost function (response R_k) is computed as a combination of pseudo-chi-squared expressions for each feature F_j within a cluster k of features:

$$R^*_j(t) = \Sigma_i [x_i(t) - F_{ji}]^2, \quad \ldots (2)$$

$$R_k(t) = \Sigma_j (1/R^*_j(t))^2 \; / \; \Sigma_j (1/R^*_j(t)). \quad \ldots (3)$$

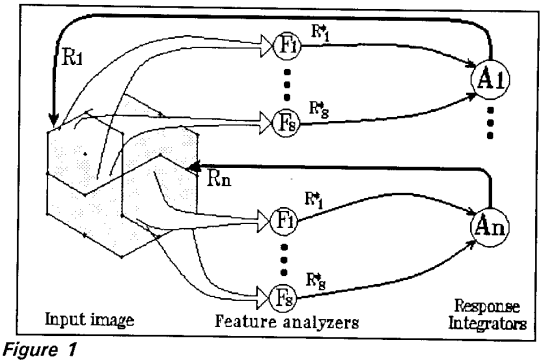

Figure 1
The neural network architecture. Each feature cluster consists of s feature analyzers F_j. The Response Integrator A_k computes the overall response R_k of the cluster k.

The goal is *maximization* of the response R_k for each feature cluster. Within each cluster, one sub-ALOPEX is running on its own data subset. Data subsets are overlapping. One iteration consists of: (a) calculation of all responses for all sub-ALOPEXes using equations (2) and (3), and (b) updating of the corresponding subset. Due to overlapping, some elements will be updated more than once in one iteration. Each upcoming sub-ALOPEX k calculates the value of a particular variable $\Delta x_i^k(t)$ for updating, using equation (1), and combines it with the existing value. The Response Integrator A_k with the lowest response should have the biggest change, and it should inhibit its neighbors the greatest. The resulting modifying equation is:

$$x_i(t) \leftarrow x_i(t) + \Delta x_i^k(t) \; / \; M^k(t) \quad \ldots (4)$$

where $M^k(t)$ is the amount of inhibition of the response integrator A_k. First we set it to "1" for all A_ks. If one of them is an order of magnitude greater than its neighbor, we increase its M by one. So, the greatest level of inhibition is 7, which happens if some response is an order of magnitude greater than all of its six neighbors. In the case of equilibrium, all response integrators will have inhibition equal to one, and this can serve as a stopping condition.

If sub-ALOPEXes in two overlapping clusters work in the same direction, this will appear as having an extra iteration, and will cause a faster convergence. If they work in opposite directions, the one who made the smaller change, will perceive changes made by the others as random noises. This actually means an increase in temperature. The temperature will be adjusted *automatically*.

3 Results

The optimizing data set is a hexagonal input image of n hexagonal elements (hexels). The templates are stored as sets of overlapping features. The goal is to get the convergence of the input

CH2936-3/90/0000-1415 $01.00

image to the one of the templates. We ran three experiments: (1) one global ALOPEX, (2) seven sub-ALOPEXes each within its own feature cluster, and (3) thirty seven sub-ALOPEXes each within its own feature cluster. The input image and the memorized templates are shown in Fig.2. Uniformly distributed noise in the range of (0-30) is added to the input image. (For a detailed discussion about selection of parameters, see [2].) The global ALOPEX needs about 3500 iterations to erode the excessive line, whereas the distributed one with seven receptive fields needs less than 1100. Very interesting is the case of smaller receptive fields, where the damage is larger than the receptive field size. The sequence of convergence in this case is shown in Fig.2. We ran several experiments, and in all cases the correct convergence appeared, and it took about 1700 iterations on the average, half of what it takes with the global ALOPEX.

4 Discussion

The method has proven to work on one particular example. It succeeded to bridge damages which were large relative to the size of the receptive fields. It is faster than the global optimization because the number of iterations will depend on the size of the damage rather than on that of data set D. It also has biological plausibility. Another benefit is that we have introduced a *stopping condition*.

This method may be applied on problems where the data set has uniform spatial density, and all local cost functions have approximately the same optimal values. The method is sufficient for some classes of problems, like the optimization of cost functions found in the problems of early vision [6]. For other classes where there is a possibility of multiple interpretation of the data, e.g. in the case of pattern recognition, additional constraints are needed. The natural constraint for pattern recognition problems is that all active feature analyzers should belong to a single template. To do this we have to introduce the organization in the first layer of neural network.

References

[1] CHON,T.-S., and E. MICHELI-TZANAKOU, "A Feature Extraction Algorithm Using Connectivity Strengths and Moment Invariants", to be published in 1990.

[2] DASEY, T.J., and E. MICHELI-TZANAKOU, "A Pattern Recognition Application of the ALOPEX Process With Hexagonal Arrays", *Proc. Int.Joint Conf.Neural Networks*, Washington DC, pp.II_119-II_125, June 1989.

[3] DURBIN, R., and D. WILLSHAW, "An Analogue Approach to the Travelling Salesman Problem Using an Elastic Net Method", *Nature*, Vol.326, pp.689-691, 16 April 1987.

[4] HOPFIELD, J.J., and D.W. TANK, "Neural Computation of Decisions in Optimization Problems", *Biological Cybernetics*, Vol.52, pp.141-152, 1985.

[5] KIRKPATRICK, S., C.D. GELATT, and M.P. VECCHI, "Optimization by Simulated Annealing", *Science*, Vol.220, No.4598, pp.671-680, 13 May 1983.

[6] POGGIO, T., V. TORRE, and C. KOCH, "Computational Vision and Regularization Theory", *Nature*, Vol.317, No.6035, pp.314-319, 26 September 1985.

[7] TAKEFUJI, Y., and H. SZU, "Design of Parallel Distributed Cauchy Machines", *Proc. Int.Joint Conf.Neural Networks*, Washington DC, pp.I_529-I_532, June 1989.

[8] TZANAKOU, E., R. MICHALAK, and E. HARTH, "The Alopex Process: Visual Receptive Fields by Response Feedback", *Biological Cybernetics*, Vol.35, pp.161-174, 1979.

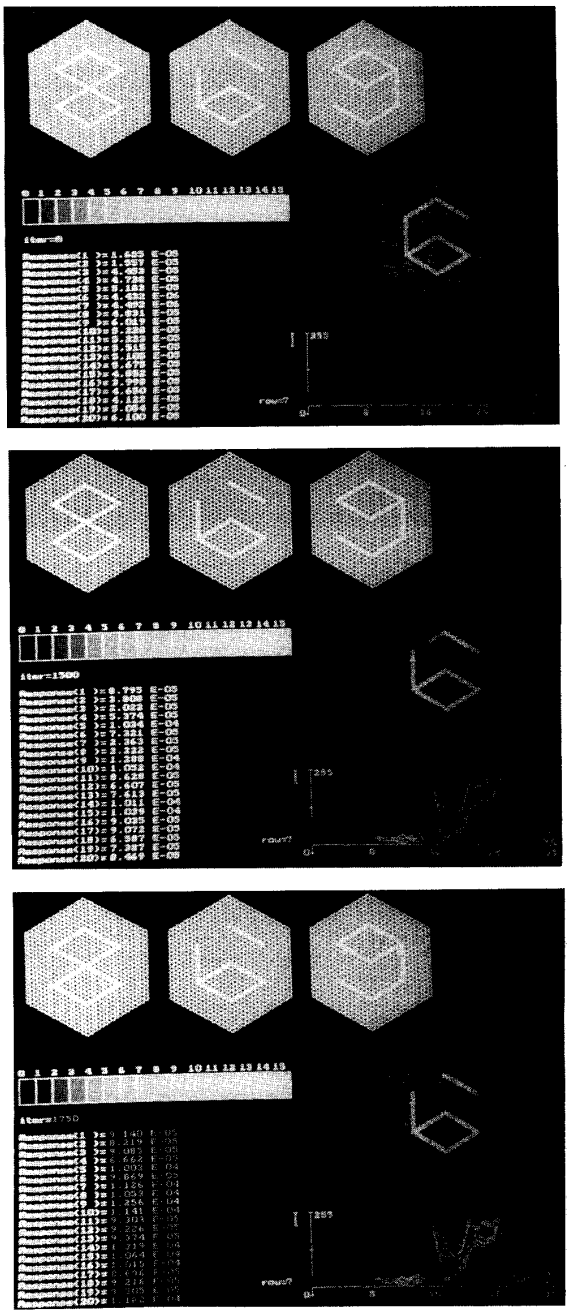

Figure 2
The sequence of convergence (0, 1500, and 1750 iterations) for the case of 37 receptive fields. The responses are shown in the left lower corner. Parameters used are: $\gamma=0.7$, and $\sigma=1.05$ (scaling factor and standard deviation of the Gaussian random number generator respectively).

ACKNOWLEDGEMENT. Research supported by NSF grant EET-8721327.