REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0138

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	18 Dec 96	FINAL TECH RPT, 15 AUG 92 to 30 JUN 96

4. TITLE AND SUBTITLE

Mathematical Methods for the Implementation of Neural Networks

5. FUNDING NUMBERS

F49620-92-J-0465

6. AUTHOR(S)

Prof. Eric Mjolsness

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Computer Science
Yale University
12 Prospect Place
New Haven CT 06511-3516

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NM
110 Duncan Avenue Suite B115
Bolling AFB DC 20332-8050

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

19980210 060

13. ABSTRACT *(Maximum 200 words)*

We present a novel optimizing network architecture with applications in vision, learning, pattern recognition and combinatorial optimization. This architecture is constructed by combining the following techniques: (i) deterministic annealing, (ii) self-amplification, (iii) algebraic transformations, (iv) clocked objectives and (v) soft assign. Deterministic annealing in conjunction with self-amplification avoids poor local minima and ensures that a vertex of the hypercube is reached. Algebraic transformations and clocked objectives help partition the relaxation into distinct phases. The problems considered have doubly stochastic matrix constraints or minor variations thereof. We introduce a new technique, soft assign, which is used to satisfy this constraint. Experimental results on different problems are presented and discussed.

14. SUBJECT TERMS

Combinatorial optimization, deterministic annealing

15. NUMBER OF PAGES 7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclsified	Unclassified	Unclassified	UL

AFOSR Final Report:
Mathematical Methods for the Implementation of Neural Networks
(Grant # F49620-92-J-0465)

Eric Mjolsness
Department of Computer Science and Engineering
YALE UNIVERSITY
NEW HAVEN CT 06511

December 18, 1996

1 Introduction

The AFOSR grant reported on here ran originally for three years, and then had a no-cost extension to the fourth year ending in mid-summer 1996 associated with the PI's move from Yale to UCSD. During that time numerous publications and a PhD thesis were produced. Most of this material is provided as an appendix to this report, to allow readers to reproduce any of the results reported here. The appendix may be consulted for a wealth of details down to the implementations of our algorithms; here we just provide an overview of work accomplished.

The research centered on the development of new and potentially automatable methods for making the transition from constrained optimization problems (which formalize applications in machine learning, vision, and many other domains) to improved (and usually neuromorphic) algorithms for solving such constrained optimization problems. The work divides naturally into the mathematical methods side and the applications side.

2 Optimization Methods

A major theme of the optimization work is to invent improved nonlinear optimization methods which can be introduced semi-automatically or automatically by means of algebraic transformations [1] on the objective functions and constraints which formulate the problem. This theme appears in a number of novel optimization methods described below. Two of these nonlinear optimization methods have been scaled up into the range of solving for around one million unknown variables.

In mathematical optimization methods, one major innovation was the "soft-assign" approach to solving combinatorial optimization problems involving unknown permuation matrices (such as quadratic assignment and graph-matching problems) which arise in computer vision, scheduling and elsewhere. The introduction of this method is amenable to automation using the "angle bracket" notation for symbolically expressing a set of algorithm phases in which different subsets of the variables are optimized. This work is detailed in a journal article [2] and a proof of convergence is forthcoming in a conference paper [3]. The method has been tested on problems at the intersection of machine vision and learning, with on the order of a million unknown discrete variables.

A second major effort was the integration of our previous nonlinear multiscale optimization algorithm for relaxation-based neural networks [4] with three other key methods: a novel "focus of attention" method for optimizing the most important set of variables at each step of the algorithm, and (more conventionally) a trust region method for robust local optimizations and a "deterministic annealing" continuation method for avoiding spurious local minima. The focus-of-attention method, like the nonlinear multiscale acceleration method, can be incorporated symbolically since the objective function for choosing the focus of attention is calculated from derivatives of the original objective function. All four methods were synthesized and shown to advantage on large-scale problems (up to about 800,000 variables) in a PhD thesis [5] and in a preceding conference paper [6] and SIAM talk [7].

Another direction for accelerating the convergence of relaxation (optimization) neural networks was demonstrated in [8] in which boundary layer methods were adapted to this class of problems.

The combinatorial optimization problems addressed efficiently by the soft-assign algorithm are in general NP-complete, so we cannot reasonably hope to invent the final algorithm. But we can explore the tradeoff between speed, algorithm size, and the quality of the local minima reached measured against the (possibly unknown) global minimum. A slower algorithm which however solves harder graph matching problems than soft-assign is introduced in [9], the journal publication of a method first presented at a neural network conference [10]. Again, the method is derived by means of an algebraic (symbolic) transformation [1] of the original graph-matching objective function.

The parallel implementation of relaxation-based neural networks is addressed from the point of view of algebraic transformations in [11], in which it is shown how to algebraically introduce extra variables at the boundaries of a partition of a large optimization problem into smaller interacting subproblems which are assigned to multiple processors in a network of workstations. The extra variables take account of communication delays between processors. Good speedup is demonstrated for large problems.

Finally a theoretical framework for dynamical systems which optimize large-scale objective functions is developed in [12], which shows how alternative dynamics such as attention mechanisms and virtual variables (virtual neurons) can be introduced using a modified Lagrangian formulation of the dynamics (not just the optimization problem). This is similar to the approach to dynamics taken in fundamental physics, except that our dynamics are dissipative. This paper also addresses the question of finding the computationally fastest (implementable) dynamical system for optimizing a given objective function, from within a family of alternative dynamical systems.

2.1 Selected Abstracts on Optimization Methods

In this subsection we reproduce the abstracts of most of the references in the Optimization Methods, to provide a second, more detailed level of overview for the work included in the Appendix. Each title cites the corresponding full paper (see the References section) and the abstracts also appear in the order in which they were introduced.

A novel optimizing network architecture with applications [2]
Anand Rangarajan, Steven Gold, Eric Mjolsness

Abstract

We present a novel optimizing network architecture with applications in vision, learning, pattern recognition and combinatorial optimization. This architecture is constructed by combining the following techniques: (i) deterministic annealing, (ii) self-amplification, (iii) algebraic transformations, (iv) clocked objectives and (v) softassign. Deterministic annealing in conjunction with self-amplification

avoids poor local minima and ensures that a vertex of the hypercube is reached. Algebraic transformations and clocked objectives help partition the relaxation into distinct phases. The problems considered have doubly stochastic matrix constraints or minor variations thereof. We introduce a new technique, softassign, which is used to satisfy this constraint. Experimental results on different problems are presented and discussed.

A convergence proof for the softassign quadratic assignment algorithm [3]
Anand Rangarajan, Alan Yuille, Steven Gold, and Eric Mjolsness

Abstract

The softassign quadratic assignment algorithm has recently emerged as an effective strategy for a variety of optimization problems in pattern recognition and combinatorial optimization. While the effectiveness of the algorithm was demonstrated in thousands of simulations, there was no known proof of convergence. Here, we provide a proof of convergence for the most general form of the algorithm.

A Multiscale Attentional Framework for Relaxation Neural Networks [6]
Dimitris I. Tsioutsias and Eric Mjolsness

Abstract

We investigate the optimization of neural networks governed by general objective functions. Practical formulations of such objectives are notoriously difficult to solve; a common problem is the poor local extrema that result by any of the applied methods. In this paper, a novel framework is introduced for the solution of large-scale optimization problems. It assumes little about the objective function and can be applied to general nonlinear, non-convex functions; objectives in thousand of variables are thus efficiently minimized by a combination of techniques - deterministic annealing, multiscale optimization, attention mechanisms and trust region optimization methods.

A Lagrangian Relaxation Network for Graph Matching [9]
Anand Rangarajan and Eric Mjolsness

Abstract

A Lagrangian relaxation network for graph matching is presented. The problem is formulated as follows: given graphs G and g, find a *permutation matrix* M that brings the two sets of vertices into correspondence. Permutation matrix constraints are formulated in the framework of *deterministic annealing*. Our approach is similar to a *Lagrangian decomposition* approach in that the row and column constraints are satisfied separately with Lagrange multipliers used to equate the two "solutions." Lagrange parameters also express the graph matching constraint. Due to the unavoidable symmetries involved in graph matching (resulting in multiple global minima), we add a *self-amplification* term in order to obtain a permutation matrix. With the application of a fixpoint preserving algebraic transformation to both the distance measure and the self-amplification terms, we obtain a Lagrangian relaxation network. The network performs minimization with respect to the Lagrange parameters and maximization with respect to the match matrix variables. Simulation results are shown on 100 node random graphs and for a wide range of connectivities.

Optimization Dynamics for Partitioned Neural Networks [11]
Dimitris I. Tsioutsias and Eric Mjolsness

Abstract

Given a relaxation-based neural network and a desired *partition* of the neurons in the network into modules with relatively slow communication between modules, we investigate relaxation dynamics for the resulting partitioned neural network. In particular, we show how the slow inter-module communication channels can be modeled by means of certain transformations of the original objective function which introduce new state variables for the inter-module communication links. We report on a parallel implementation of the resulting relaxation dynamics, for a two-dimensional image segmentation network, using a network of workstations. Experiments demonstrate a functional and efficient parallelization of this neural network algorithm. We also discuss implications for analog hardware implementations of relaxation networks.

A Lagrangian Approach to Fixed Points [13]
Eric Mjolsness and Willard L. Miranker

Abstract

We present a new way to derive dissipative, optimizing dynamics from the Lagrangian formulation of mechanics. It can be used to obtain both standard and novel neural net dynamics for optimization problems. To demonstrate this we derive standard descent dynamics as well as nonstandard variants that introduce a computational attention mechanism.

Greedy Lagrangians for Neural Networks: Three Levels of Optimization in Relaxation Dynamics [12]
Eric Mjolsness and Willard L. Miranker

Abstract

We expand the mathematical apparatus for relaxation networks, which conventionally consists of an objective function E and a dynamics given by a system of differential equations whose trajectories diminish E. Instead we (1) retain the objective function E, in a standard neural network form, as the measure of the network's computational *functionality*; (2) derive the dynamics from a Lagrangian function L which depends on both E and a measure of computational *cost*; and (3) tune the form of the Lagrangian according to a meta-objective $\mathcal{M}$ which may involve measuring cost and functionality over many runs of the network. The essential new features are the Lagrangian, which specifies an objective function that depends on the neural network's state over all times (analogous to Lagrangians which play a similar fundamental role in physics), and its associated *greedy functional derivative* from which neural-net relaxation dynamics can be derived.

The combination of Lagrangian and meta-objective suffice to derive and provide an interpretation for *clocked objective functions*, a useful notation for algebraically formulating and designing neural network applications, possibly with the assistance of symbolic computation. Clocked objectives thus generalize the original static objective function E as a practical neural network specification language.

With these methods we are able to analyze the approximate optimality of Hopfield/Grossberg dynamics, the generic emergence of sub-problems involving learning and scheduling as aspects of relaxation-based neural computation, the integration of relaxation-based and feed-forward neural networks, and the control of computational *attention mechanisms* using priority queues, coarse-scale blocks of neurons, default-valued neurons, and other special-case optimization algorithms.

3 Applications

Most but not all of the applications with which we experimented were taken from computer vision and learning, using either dense images or sparse image feature sets as data. The multiscale attention mechanism [5, 6] was tested on large image segmentation problems as well as on more abstract graph-partition problems.

A major class of applications in computer vision is related to *correspondence problems* between two sparse image feature sets, i.e. finding which feature if any in one image corresponds to which feature in another image, and deriving the consequences of such identifications. We solve correspondence problems under a wide variety of noise conditions in [14] using the soft-assign optimization algorithm described in [2], and use that capability to *learn* new object models (themselves sparse feature sets) from unlabelled data in [15, 16].

The model-learning experiments and the region segmentation experiments were each extended to large-scale global nonlinear optimization problems, on the order of a million variables.

3.1 Selected Abstracts on Applications

In this subsection we reproduce the abstracts of most of the references in the Applications section, to provide a second, more detailed level of overview for the work included in the Appendix. Each title cites the corresponding full paper (see the References section) and the abstracts also appear in the order in which they were introduced.

New Algorithms for 2D and 3D Point Matching:
Pose Estimation and Correspondence [14]

Steven Gold, Anand Rangarajan, Chien-Ping Lu, Suguna Pappu, and Eric Mjolsness

Abstract

A fundamental open problem in computer vision—determining pose and correspondence between two sets of points in space—is solved with a novel, fast, robust and easily implementable algorithm. The technique works on noisy 2D or 3D point sets that may be of unequal sizes and may differ by non-rigid transformations. Using a combination of optimization techniques such as deterministic annealing and the *softassign*, which have recently emerged out of the recurrent neural network/statistical physics framework, analog objective functions describing the problems are minimized. Over thirty thousand experiments, on randomly generated points sets with varying amounts of noise and missing and spurious points, and on hand-written character sets demonstrate the robustness of the algorithm.

Learning with Preknowledge:
Clustering with Point and Graph Matching Distance Measures [15]

Steven Gold, Anand Rangarajan and Eric Mjolsness

Abstract

Prior knowledge constraints are imposed upon a learning problem in the form of distance measures. Prototypical 2-D point sets and graphs are learned by clustering with

point matching and graph matching distance measures. The point matching distance measure is invariant under affine transformations - translation, rotation, scale and shear - and permutations. It operates between noisy images with missing and spurious points. The graph matching distance measure operates on weighted graphs and is invariant under permutations. Learning is formulated as an optimization problem. Large objectives so formulated ($\sim$ million variables) are efficiently minimized using a combination of optimization techniques - algebraic transformations, projection methods, clocked objectives, and deterministic annealing.

Clustering with a Domain-Specific Distance Measure [17]
Steven Gold, Eric Mjolsness and Anand Rangarajan

Abstract

With a point matching distance measure which is invariant under translation, rotation and permutation, we learn 2-D point-set objects, by clustering noisy point-set images. Unlike traditional clustering methods which use distance measures that operate on feature vectors – a representation common to most problem domains – this object-based clustering technique employs a distance measure specific to a type of object within a problem domain. Formulating the clustering problem as two nested objective functions, we derive optimization dynamics similar to the Expectation-Maximization algorithm used in mixture models.

4 Conclusion

We have described the AFOSR-funded work at three levels of detail. First, we provided a broad overview of (a) neuromorphic mathematical optimization methods amenable to algebraic manipulation, and (b) a few of their applications. Second, we included the abstracts of most of the resulting conference and journal papers to add a little more technical detail. Finally in the Appendix, for complete ~~scientific~~ reproducibility, we include the actual papers on these topics ~~along with~~ a PhD dissertation. Withdrawn

References

[1] E. Mjolsness and C. Garrett. Algebraic transformations of objective functions. *Neural Networks*, 3:651–669, 1990.

[2] A. Rangarajan, S. Gold, and E. Mjolsness. A novel optimizing network architecture with applications. *Neural Computation*, 8(5):1041–1060, 1996.

[3] A. Rangarajan, A. Yuille, S. Gold, and E. Mjolsness. A convergence proof for the softassign quadratic assignment algorithm. In *Advances in Neural Information Processing Systems 9*. MIT Press, 1997. Forthcoming.

[4] Eric Mjolsness, Charles D. Garrett, and Willard L. Miranker. Multiscale optimization in neural nets. *IEEE Transactions on Neural Networks*, 2(2), March 1991.

[5] Dimitris I. Tsioutsias. *Multiscale Attention as a Globally Convergent Framework for Large-Scale Nonlinear Optimization.* PhD thesis, Yale University Department of Electrical Engineering, 1997. Forthcoming.

[6] Dimitris I. Tsioutsias and Eric Mjolsness. A multiscale attentional framework for relaxation neural networks. In *Advances in Neural Information Processing Systems 8*. MIT Press, 1996.

[7] Dimitris I. Tsioutsias and Eric Mjolsness. Large-scale optimization by a multiscale-attentional scheme. Talk at the Fifth SIAM Conference on Optimization, Society for Industrial and Applied Mathematics, Victoria BC, May 22 1996.

[8] Willard L. Miranker and Eric Mjolsness. Accelerating neural net dynamics by boundary layer methods. *Applied Mathematics Letters*, 6(3):85–89, 1993.

[9] A. Rangarajan and E. Mjolsness. A Lagrangian relaxation network for graph matching. *IEEE Trans. on Neural Networks*, 7(6), 1996.

[10] A. Rangarajan and E. Mjolsness. A Lagrangian relaxation network for graph matching. In *IEEE Intl. Conf. on Neural Networks (ICNN)*, volume 7, pages 4629–4634. IEEE Press, 1994.

[11] Dimitris I. Tsioutsias and Eric Mjolsness. Optimization dynamics for partitioned neural networks. *International Journal of Neural Systems*, December 1994.

[12] E. Mjolsness and W. Miranker. Greedy Lagrangians for neural networks: three levels of optimization in relaxation dynamics. Technical Report YALEU/DCS/TR-945, Department of Computer Science, Yale University, Jan 1993.

[13] Eric Mjolsness and Willard L. Miranker. A lagrangian approach to fixed points. In Lippmann, Moody, and Touretsky, editors, *Advances in Neural Information Processing Systems 3*. Morgan-Kaufman, 1991.

[14] S. Gold, C. P. Lu, A. Rangarajan, S. Pappu, and E. Mjolsness. New algorithms for 2-D and 3-D point matching: pose estimation and correspondence. *Pattern Recognition*, (in press), 1996.

[15] S. Gold, A. Rangarajan, and E. Mjolsness. Learning with preknowledge: clustering with point and graph matching distance measures. *Neural Computation*, 8(4):787–804, 1996.

[16] S. Gold, A. Rangarajan, and E. Mjolsness. Learning with preknowledge: clustering with point and graph matching distance measures. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, *Advances in Neural Information Processing Systems 7*. MIT Press, Cambridge, MA, 1995.

[17] S. Gold, E. Mjolsness, and A. Rangarajan. Clustering with a domain specific distance measure. In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems 6*, pages 96–103. Morgan Kaufmann, San Francisco, CA, 1994.