RESEARCH STATEMENT

Hillol Kargupta

1 Ph.D. Dissertation

Title: SEARCH, Polynomial Complexity, And The Fast Messy Genetic Algorithm

Abstract: Blackbox optimization (BBO)—optimization in presence of limited knowledge about the objective function—has recently enjoyed a large increase in interest because of the demand from the practitioners. This has triggered a race for new high performance algorithms for solving large, difficult problems. Simulated annealing, genetic algorithms, tabu search are some examples. Unfortunately, each of these algorithms is creating a separate field in itself and their use in practice is often guided by personal discretion rather than scientific reasons. The primary reason behind this confusing situation is the lack of any comprehensive understanding about blackbox search. This dissertation takes a step toward clearing some of the confusion.

The main objectives of this dissertation are:

1. present SEARCH (Search Envisioned As Relation & Class Hierarchizing)—an alternate perspective of blackbox optimization and its quantitative analysis that lays the foundation essential for transcending the limits of random enumerative search;

2. design and testing of the fast messy genetic algorithm.

SEARCH is a general framework for understanding blackbox optimization in terms of relations, classes and ordering. The primary motivation comes from the observation that sampling in blackbox optimization is essentially an inductive process and in absence of any relation among the members of the search space, induction is no better than enumeration. The foundation of SEARCH is laid on a decomposition of BBO into relation, class, and sample spaces. An ordinal, probabilistic, and approximate framework is developed on this foundation to identify the fundamental principles in blackbox optimization, essential for transcending the limits of random enumerative search. Bounds on success probability and sample complexity are derived. I explicitly consider specific blackbox algorithms like simulated annealing, genetic algorithms and demonstrate that the fundamental computations in all of them can be captured using SEARCH. SEARCH also offers an alternate perspective of natural evolution that establishes the computational role of gene expression (DNA→RNA→Protein) in evolution. This model of evolutionary computation hypothesizes a possible mapping of the decomposition in relation, class, and sample spaces of SEARCH into the transcriptional regulatory mechanisms, proteins, and DNA respectively.

The second part of this dissertation starts by noting the limitations of simple GAs, which fail to properly search for relations and makes decision making very noisy by combining relation, class, and the sample spaces. Messy genetic algorithms (Goldberg, Korb, & Deb, 1989) are a rare class of algorithms that emphasize the search for relations. Despite this strength of messy GAs, they lacked complete benefits of implicit parallelism (Holland, 1975). The fast messy GA initiated by Goldberg, Deb, Kargupta, and Harik (1993) introduced some of the benefits of implicit parallelism in messy GA without sacrificing its other strengths very much. This dissertation investigates fast messy GAs and presents test results to demonstrate its performance for order-k delineable problems.

2 Other Research:

I have done the following works during my graduate life and they are outside the scope of my dissertation.

- *Deceptive representation*: A deceptive representation is the one in which detecting appropriate relations to solve the optimization problem is difficult. I showed that the class of partially deceptive problems is NP-Turing complete. I also showed that whether or not a representation is deceptive representation is NP hard.

- *Crosstalk*: Crosstalk problem occurs in blackbox optimization when evaluation of multiple relations interfere with each other. The accuracy and reliability of the decision making decreases as a result. I analyzed crosstalk using the classical multiple armed framework and developed a Bayesian approach to quantify crosstalk in genetic algorithms.

- Application of Fast Messy GA for *scan-to-scan missile correlation* problem (sponsored by United States Army, Contract DASG60-90-C0153). This is a pattern recognition problem involving the determination of correlations among multiple missile blips in consequent image frames.

- Spatio-temporal sequence prediction using reaction-diffusion systems: Natural reaction-diffusion systems show very interesting pattern formation process. We observed that the limb regeneration process of cockroach could be useful for sequence prediction problem. We came up with a system that performed impressively for Bio-Medical signal processing problems (Kargupta, Ray, 94) (Ray, Kargupta, 95).
- Knowledge based systems: An expert system was developed for solving partial differential equations, which was a module of a heat exchanger design system. This was my master's thesis (Sahay, Kargupta, 91)
- Decision support systems: A virtual decision support system for army installation maintenance was developed (Fancois, Subick, Kargupta, Govindan, 94). This knowledge based system has a discrete event simulation environment and an optimization module based on genetic algorithms which helps the user allocating resources in a near optimal way.
- Parallel Genetic Algorithms: GAs are inherently very parallel in nature; yet very little theoretical work has so far been done in this direction. We came up with an analytical recommendation for population sizing for parallel GAs with both no and partial migration (Goldberg, Kargupta, Horn and Catupaz, 94).
- Statistical mechanics and genetic algorithms: Many interesting connections between natural evolution and statistical mechanics have been exposed in the past. I did some work on quantifying the notion of *temperature* for population based search techniques using formal information theory (Kargupta, 93) and showed that *Boltzmann distribution* can be formed in a drift-diffusion GA (Kargupta, 92).
- System identification: My background interest led me to do some applications in nonlinear system identification (Kargupta, Smith, 91).

3 Future Research Plan:

The following sections describe my immediate future research directions:

3.1 Development of Gene Expression Messy GA

The SEARCH framework developed in my dissertation set the stage for designing BBO algorithms in a constructive way. Gene Expression Messy GA (GEMGA) is an example of that. Unlike most of the existing BBO algorithms GEMGA directly search for relations. It first detects the appropriate relations among the members of the search space and then use those relations to find good solutions. GEMGA is an $O(2^k(\ell+k))$ sample complexity algorithm for the class of order-k delineable problems (Kargupta, 1995a) (problems that can be solved by considering no higher than order-k relations) of length ℓ. GEMGA is designed based on alternate perspective of natural evolution proposed by the SEARCH framework that emphasizes the role of gene expression. GEMGA uses the *transcription* operator to search for relations. GEMGA has been tested on very large multimodal problems with millions of local optima. It is currently under further rigorous testing against different classes of problems.

3.2 Practical applications of Gene Expression Messy GA

GEMGA awaits immediate applications in different fields. Some of the directions, currently being explored, are,

- **Molecular conformation problems:** This algorithm is used to find the global energy minimum solution for 19, 37 and 61 atom artificially constructed molecules, developed at Sandia National Laboratory. The number of function evaluations is much less compared to the results reported earlier in the literature. This work is currently under progress.
- **Computer scheduling problem:** This work is currently investigated for the supercomputing facility at Los Alamos National Laboratory.
- **Data mining:** Possible collaborative efforts with National Center for Supercomputing Applications, Urbana, IL and its corporate sponsors.
- **VLSI design.**

3.3 Autonomous agents and robotics

My background in mechanical engineering motivated me to work in an area that combines my current research with my knowledge in mechanical engineering. Application of GEMGA for cooperative problem solving using autonomous robot agents is one way to cater my desire for such synergetics. I have started working with a group of people with controls background in Los Alamos National Laboratory in this area. The possible applications that we are currently looking at are detection of underground mines, and hazardeous material handling. I am highly interested in pursuing this effort in future.

3.4 Extension of the SEARCH framework

The SEARCH framework addressed unconstrained global blackbox optimization. This lead to several future directions on the theoretical front. Some of them are,

- extension to constrained optimization problem.
- incorporating the SEARCH framework into verified interval optimization.

3.5 Universal computation, evolution and many possibilities

Ever since Turing and Church's model of computation brought the scientific community to a new era, many different models of computation have been proposed. Although they have been proved to be equivalent to the Turing machine model, some of them may provide useful alternatives for new generation computer architecture design. One such example is the paradigm of *reversible computation*, which considers computation as a physical process, subjected to the physical laws, such as laws of thermodynamics. Reversible computation provides a framework for performing nondissipative computation. Experimental benefits (low heat dissipating chips) of nondissipative computation has been reported recently from both industry and academia, such as MIT (Fredkin, Toffoli, 82), XEROX Palo Alto (Merkle, 92), IBM (Landauer, 1992), Fujitsu (Kiehl, 94). Simulations of physical systems through particle based computation has also been found useful in such alternate architectures of computations (cellular automata machines).

Developing a logical algebra for the evolutionary operators may also lead to alternate ways to perform universal computation with careful attention to laws of thermodynamics. For example, a universal NOR gate can be easily constructed using a single gene primitive crossover (Kargupta, 95f), which is in fact looks exactly same as Fredkin gate (Fredkin, Toffoli, 82). Since evolutionary processes are fundamental in almost every natural systems, such models of logical computations are likely to provide a natural way to simulate evolving systems. I am highly interested in developing a logical algebra for evolutionary mechanism and then pursuing a possible implementation using charged-coupled devices.

4 References

For the references with my last name (Kargupta) in the author list, please see the publication list, enclosed with the resume.

1. Fredkin, E. and Toffoli, T. (1982). Conservative logic. International Journal of Theoretical Physics. vol. 21, Nos. 3/4, pp 219–253.

2. Goldberg, D. E., Korb, B., & Deb, K. (1989). *Messy genetic algorithms: Motivation, analysis, and first results.* Complex Systems. pp493-530.

3. Holland, J. (1975). *Adaptation in Natural and Artificial Systems*, The University of Michigan press, Ann Arbor, Michigan.

4. Kiehl, R. (1994). Research toward nano-electronic computing technologies in Japan. Proceedings of the workshop on physics and computation. IEEE Computer Society Press. pp 1-4.

5. Landauer, R. (1992). Information is physical. Proceedings of the workshop on physics and computation. IEEE Computer Society Press. pp 1-4.

6. Merkle, R. (1992). Towards practical reversible logic. *Proceedings of the Workshop on Physics and Computation*, IEEE Computer Society Press, pp. 227–228.