Missouri Estimation of Distribution Algorithms Laboratory

Enhancing Efficiency of Hierarchical BOA via Distance-Based Model Restrictions

Mark Hauschild and Martin Pelikan

MEDAL Report No. 2008007

April 2008

Abstract

This paper analyzes the effects of restricting probabilistic models in the hierarchical Bayesian optimization algorithm (hBOA) by defining a distance metric over variables and disallowing dependencies between variables at distances greater than a given threshold. We argue that by using prior problem-specific knowledge, it is often possible to develop a distance metric that closely corresponds to the strength of interactions between variables. This distance metric can then be used to speed up model building in hBOA. Three test problems are considered: 3D Ising spin glasses, random additively decomposable problems, and the minimum vertex cover.

Keywords

Hierarchical BOA, efficiency enhancement, learning from experience, probabilistic model, model structure, model complexity, estimation of distribution algorithms.

Missouri Estimation of Distribution Algorithms Laboratory (MEDAL)
Department of Mathematics and Computer Science
University of Missouri–St. Louis
One University Blvd., St. Louis, MO 63121
E-mail: medal@cs.umsl.edu
WWW: http://medal.cs.umsl.edu/

Enhancing Efficiency of Hierarchical BOA via Distance-Based Model Restrictions

Mark Hauschild
Missouri Estimation of Distribution Algorithms Laboratory (MEDAL)
Dept. of Math and Computer Science, 320 CCB
University of Missouri at St. Louis
One University Blvd., St. Louis, MO 63121
mwh308@umsl.edu

Martin Pelikan
Missouri Estimation of Distribution Algorithms Laboratory (MEDAL)
Dept. of Math and Computer Science, 320 CCB
University of Missouri at St. Louis
One University Blvd., St. Louis, MO 63121
pelikan@cs.umsl.edu

Abstract

This paper analyzes the effects of restricting probabilistic models in the hierarchical Bayesian optimization algorithm (hBOA) by defining a distance metric over variables and disallowing dependencies between variables at distances greater than a given threshold. We argue that by using prior problem-specific knowledge, it is often possible to develop a distance metric that closely corresponds to the strength of interactions between variables. This distance metric can then be used to speed up model building in hBOA. Three test problems are considered: 3D Ising spin glasses, random additively decomposable problems, and the minimum vertex cover.

Keywords: Hierarchical BOA, efficiency enhancement, learning from experience, probabilistic model, model structure, model complexity, estimation of distribution algorithms.

1 Introduction

The hierarchical Bayesian Optimization Algorithm (hBOA) (Pelikan & Goldberg, 2001; Pelikan, 2005) has been shown to solve a large range of problems scalably and robustly. However, being able to solve a problem in low-order polynomial time is not always enough. As problem size and difficulty increases, the computational resources necessary could still make the problem intractable in practice. That is why it is important to design efficiency enhancement techniques (Sastry, Goldberg, & Pelikan, 2001; Pelikan & Sastry, 2004; Lima, Pelikan, Sastry, Butz, Goldberg, & Lobo, 2006; Pelikan, Sastry, & Goldberg, 2006; Sastry, Pelikan, & Goldberg, 2006) which can further improve the efficiency of hBOA and other estimation of distribution algorithms (EDAs) (Baluja, 1994; Mühlenbein & Paaß, 1996; Larranaga & Lozano, 2001; Pelikan, 2005).

One approach to speeding up EDAs is to use prior knowledge to restrict or bias model building (Schwarz & Ocenasek, 2000; Mühlenbein & Mahnig, 2002; Baluja, 2006; Santana, Larrañaga,

& Lozano, 2007). One way to do this is to define a distance metric on variables such that the variables that are close to each other with respect to this metric are expected to influence each other more strongly. If this metric accurately reflects the problem structure, one can use it to speed up hBOA and other EDAs with complex models by disallowing dependencies between variables at a distance above a given threshold (Mühlenbein & Mahnig, 2002; Baluja, 2006; Santana, Larrañaga, & Lozano, 2007). This should improve algorithm performance by simplifying model building and increasing model accuracy.

In this paper we study the effects of such model restrictions on three important classes of problems: 3D spin glasses, random additively decomposable problems, and minimum-vertex cover. First, a distance metric is constructed for each problem based on the structure of the objective function. The metric is then used to restrict models in hBOA, and hBOA performance is tested for various values of the threshold. The results show that when hBOA is given a reasonable threshold, model restrictions lead to substantial speedups.

The paper is organized as follows. Section 2 discusses prior work on biasing model building in EDAs. Section 3 outlines hBOA. Section 4 discusses the design of good distance metrics and the expected benefits of using an appropriate distance metric to restrict model structure. Section 5 describes test problems and the distance metrics for each problem. Section 6 presents experimental results. Finally, section 7 summarizes and concludes the paper.

2 Previous work on biasing model building in EDAs

There are two main approaches to biasing model building in EDAs: (1) Impose *soft restrictions* by biasing the scoring metric to prefer models that closely correspond to the problem structure (Schwarz & Ocenasek, 2000; Hauschild, Pelikan, Sastry, & Goldberg, 2008) or (2) impose *hard restrictions* by strictly disallowing some dependencies (Mühlenbein & Mahnig, 2002; Baluja, 2006; Santana, Larrañaga, & Lozano, 2007; Hauschild, Pelikan, Sastry, & Goldberg, 2008).

Schwarz & Ocenasek (Schwarz & Ocenasek, 2000) proposed the use of prior probabilities of competing network structures in BOA to bias model building toward models that closely correspond to the problem structure in graph bi-partitioning. Edges between variables connected in the underlying graph were thus given preference, but no edges were strictly disallowed.

Mühlenbein & Mahnig (Mühlenbein & Mahnig, 2002) also considered graph bi-partitioning but they used a hard restriction to only allow connections between nodes connected in the underlying graph. Baluja (Baluja, 2006) proposed the use of the same hard restriction in the dependency-tree EDA on graph coloring. Santana (Santana, Larrañaga, & Lozano, 2007) used the same hard restriction to speed up model building of dependency trees on a protein design problem. Hauschild et al. (Hauschild, Pelikan, Sastry, & Goldberg, 2008) proposed the use of both soft restrictions based on prior models on problems of the same structure as well as hard restrictions based on a distance metric for hBOA on the 2D spin glass and MAXSAT. This paper extends the work of Hauschild et al. (Hauschild, Pelikan, Sastry, & Goldberg, 2008) on hard restrictions based on a distance-metric by considering other important classes of problems, including two NP-complete problems (3D spin glass and minimum vertex cover).

3 Hierarchical BOA (hBOA)

Estimation of distribution algorithms (EDAs) (Baluja, 1994; Mühlenbein & Paaß, 1996; Larrañaga & Lozano, 2002; Pelikan, 2005; Pelikan, Goldberg, & Lobo, 2002) replace standard crossover and mutation operators of genetic algorithms by building an explicit probabilistic model of selected solutions and sampling the built model to generate new candidate solutions. hBOA is an EDA that uses Bayesian networks as probabilistic models and incorporates restricted tournament replacement (Harik, 1995) for effective diversity maintenance.

hBOA evolves a population of candidate solutions represented by fixed-length strings over a finite alphabet (e.g., binary strings). The initial population is generated at random according to the uniform distribution over all potential solutions. Each iteration (generation) starts by selecting promising solutions from the current population using any standard selection method of genetic and evolutionary algorithms. In this paper we use truncation selection with threshold $\tau = 50\%$. Next, hBOA builds a Bayesian network (Howard & Matheson, 1981; Pearl, 1988) with local structures (Chickering, Heckerman, & Meek, 1997; Friedman & Goldszmidt, 1999) as a model of the selected solutions using a greedy algorithm. Learning the model structure is typically the most challenging task of model building (Pelikan, 2005). New solutions are generated by sampling the built network. These are then incorporated into the original population using restricted tournament replacement (RTR) (Harik, 1995), which ensures effective diversity maintenance. We set the window size $w = \min\{n, N/20\}$, where n is the number of decision variables and N is the population size, as suggested in ref. (Pelikan, 2005). The next iteration is then executed unless some predefined termination criteria are met. For more details on learning and sampling BNs with local structures in hBOA, see refs. (Chickering, Heckerman, & Meek, 1997; Pelikan & Goldberg, 2001; Pelikan, 2005; Friedman & Goldszmidt, 1999).

Since a simple deterministic hill climber (DHC) was shown to lead to substantial speedups of hBOA on both the 3D spin glass (Pelikan, 2005) and the minimum vertex cover (Pelikan, Kalapala, & Hartmann, 2007), we incorporated DHC into hBOA for these two problems also in this study. DHC takes a candidate solution and performs one-bit changes on it that lead to the maximum improvement in fitness. DHC is terminated when no single-bit flip leads to improvement. This is done on each solution before it is evaluated.

4 Distance-Based Model Restriction

To improve model building in hBOA, only necessary dependencies should be considered. This may significantly reduce the space of potential model structures, improving the speed and accuracy of model building. Doing this is often not easy in practice, as for many problems it is difficult to identify necessary dependencies. Nonetheless, it is often possible to provide a soft measure that ranks edges according to their expected importance based on prior problem-specific knowledge. The likelihood of certain dependencies can be estimated for example from an explicit or implicit distance metric between problem variables. In many problems, such a distance metric is relatively easy to design from the structure of the objective function or previous runs of an EDA on similar problem instances.

For example, for 2D spin glasses, Hauschild et al. (Hauschild, Pelikan, Lima, & Sastry, 2007) showed that while in general it is not easy to decide what dependencies are unnecessary, dependencies are more likely to connect spins located close to each other with respect to the shortest path between these spins in the underlying 2D lattice. This fact is explored by Hauschild et al. (Hauschild, Pelikan, Sastry, & Goldberg, 2008), who examined the effects of specifying a distance threshold and disallowing dependencies between spins at a distance above the threshold. A similar distance metric was proposed and tested for MAXSAT. The results showed that with an appropriate threshold, distance-based restrictions lead to substantial speedups on both problems. This paper considers similar restrictions on three additional classes of difficult problems, two of which are NP-complete.

We expect two main benefits from restricting model structure in this way. By disallowing some dependencies, model building should become significantly faster because there are fewer dependencies to examine. Additionally, by disallowing unlikely dependencies, the model should contain fewer spurious dependencies, leading to improved model accuracy. These two factors should lead to significant speedup of hBOA as is also confirmed in section 6.

The following section describes the test problems considered in this paper and the distance metrics used to restrict models for all problems.

5 Test Problems

5.1 3D Ising Spin Glass with $\pm J$ couplings

An Ising spin glass is typically arranged on a regular 2D or 3D grid where each node i corresponds to a spin s_i and each edge $\langle i, j\rangle$ corresponds to a coupling between two spins s_i and s_j. For the classical Ising model, each spin s_i can be in one of two states: $s_i = +1$ or $s_i = -1$. Each edge has a real value $J_{i,j}$ associated with it that defines the relationship between the two connected spins. Periodic boundary conditions are used that introduce a coupling between the first and the last elements along each dimension.

Here the task is to find spin configurations C that minimize the energy for a given set of coupling constants $J_{i,j}$, defined as

$$E(C) = \sum_{\langle i,j\rangle} -s_i J_{i,j} s_j \ , \tag{1}$$

where the sum runs over all couplings $\langle i, j\rangle$. The minimum-energy configurations are called ground states. To represent spin configurations, we use binary strings of length n where ith bit defines the value of the ith spin (-1 is encoded by 0, +1 is encoded by 1). We consider random instances of the 3D $\pm J$ spin glass, where each coupling constant is set randomly to $+1$ or -1 with equal probability. The problem of finding ground states of 3D spin glasses is NP-complete (Barahona, 1982) and, due to its complex landscape, it poses a challenge for most optimization algorithms.

Based on the results on 2D spin glasses (Hauschild, Pelikan, Lima, & Sastry, 2007; Hauschild, Pelikan, Sastry, & Goldberg, 2008), we define the distance between two spins as the shortest path between these spins in the underlying 3D grid.

5.2 Random Additively Decomposable Problems

Random additively decomposable problems (rADPs) (Pelikan, Sastry, Butz, & Goldberg, 2006b; Sastry, Pelikan, & Goldberg, 2007) are a class of test problems developed to test performance of evolutionary algorithms on broad classes of decomposable problems. The input string in rADPs is partitioned into subsets of bits, with the overall fitness being the sum of the subfunctions applied to all the subsets. We denote the order of rADPs by k; that is, each subproblem contains k bits. In order to ensure that rADP instances are solvable in polynomial time, the subproblems are located in contiguous blocks of k bits and the overlap is specified by a parameter o which denotes the number of bits shared by neighbor subproblems. Similarly as in ref. (Pelikan, Sastry, Butz, & Goldberg, 2006a), the fitness for each subproblem is given by a table of 2^k values which are generated randomly from the uniform distribution over $[0, 1)$. To verify the global optimum, the dynamic-programming algorithm (Pelikan, Sastry, Butz, & Goldberg, 2006a) is used. Instances of rADPs vary in difficulty due to the differences in subfunction difficulty and the amount of overlap.

Intuitively, bits located in the same subproblem are likely to influence each more strongly than bits located in different subproblems regardless of the overlap. More generally, in the presence of overlap, we can expect that the greater the distance between the subproblems containing two bits, the more weakly the two bits influence each other (Hauschild, Pelikan, Lima, & Sastry, 2007). To design a distance metric that captures this fact, we create a graph that connects pairs of bits located in the same subproblem and define the distance between each such pair of bits as 1. The distance between any two bits is then computed as the shortest path between these bits in this graph.

5.3 Minimum Vertex Cover

The minimum vertex cover (MVC) of an undirected graph G is the smallest subset of nodes in G such that for every edge G, at least one of the two nodes this edge connects is in this subset. MVC is interesting because it is NP-complete (Karp, 1972) and is closely related to other hard graph problems. In this paper we consider MVC for random instances of $G(n,m)$ graphs (Bollobas, 2001; Weigt & Hartmann, 2001). $G(n,m)$ consists of graphs with n vertices and m edges such that $m = nc$, where $c > 0$ is a constant. To represent subsets of nodes in hBOA, we use n-bit binary strings where the ith bit is 1 if and only if the ith node is selected. A repair operator is used to ensure that each solution corresponds to a valid graph cover (Pelikan, Kalapala, & Hartmann, 2007) and the fitness of each cover is defined as the number of nodes *not* contained in this cover.

Intuitively, bits corresponding to the nodes located closer in the underlying graph can be expected to influence each other more strongly. Accordingly, we define the distance between vertices as the minimum number of edges on a path between these vertices in the graph. Pairs of vertices in unconnected components are assigned distance n.

6 Experiments

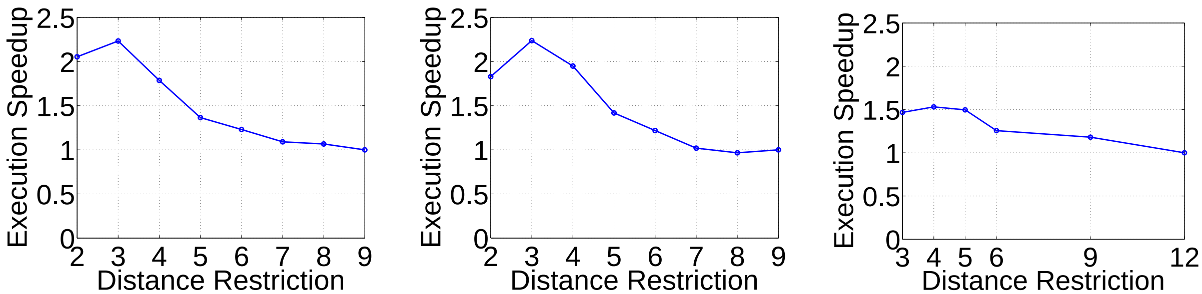
For all problem instances, bisection (Sastry, 2001; Pelikan, 2005) was used to determine the minimum population size to ensure convergence to the global optimum in 5 out of 5 independent runs, with the results averaged over the 5 runs. The number of generations was upper bounded according to hBOA scalability theory (Pelikan, Sastry, & Goldberg, 2002) by n where n is the number of bits in the problem. Each run of hBOA is terminated when the global optimum has been found (success) or when the upper bound on the number of generations has been reached without discovering the global optimum (failure).

Various values of the distance threshold parameter are considered, from the maximum observed distance to the minimum distance required to solve all problem instances in a reasonable time. The results for certain thresholds are excluded since the restriction was found too severe, requiring extremely large population sizes, $N \geq 10^5$.

6.1 Results on 3D spin glasses

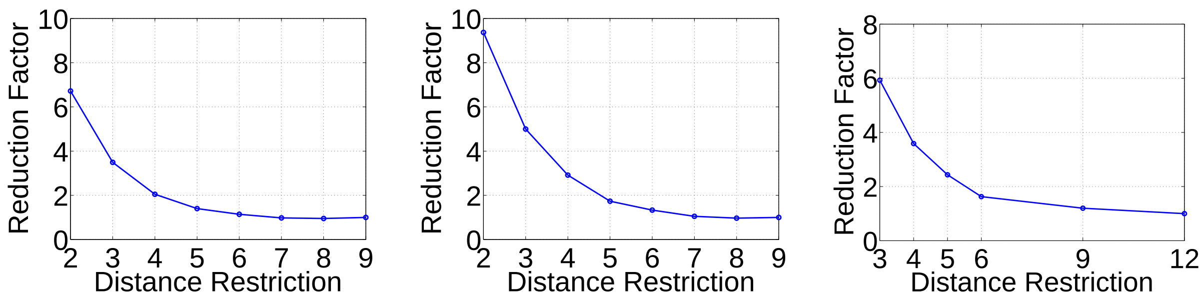
To examine the speedups obtained with distance-based model restrictions in 3D spin glasses, we considered three problem sizes: $6 \times 6 \times 6$, $7 \times 7 \times 7$ and $8 \times 8 \times 8$. Since spin glass instances vary in difficulty, we considered 1000 different instances for the two smaller sizes and 100 different instances for the largest size.

Figure 1 shows the execution-time speedup for various distance thresholds for all three problem sizes. The results show that the model restriction yields speedups of 1.5 two 2.2. We also see



(a) $6 \times 6 \times 6$ (b) $7 \times 7 \times 7$ (c) $8 \times 8 \times 8$

Figure 1: Execution time speedup by distance restriction on 3D spin glass.



(a) $6 \times 6 \times 6$ (b) $7 \times 7 \times 7$ (c) $8 \times 8 \times 8$

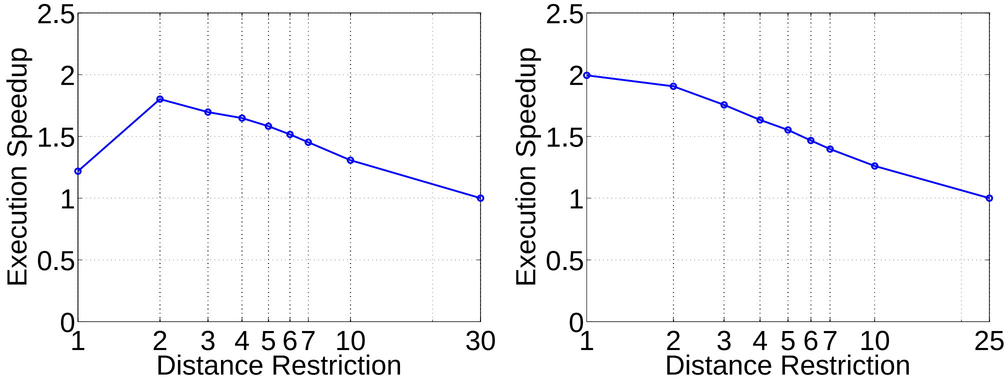
Figure 2: Reduction in the number of bits examined in model building on 3D spin glass.

that, as expected, as the problem becomes larger, dependencies at larger distances are expected to be important. For the two smaller problems, the best threshold is 3, whereas for the largest problem the best thresholds are 4 and 5. Furthermore, in the two smaller problems, distance of at least 2 must be considered for efficient performance, whereas for the largest problem we must consider distances of at least 3. The results also show that while the best speedups obtained for the two smaller problems are almost identical, the best speedup obtained on the largest instance is smaller. The reason for this may be that only 100 problem instances could be considered for the largest problem due to its high complexity, and the results are affected by the specific selection of instances.

Figure 2 shows the reduction in the number of bits examined during the entire model-building procedure. The results show a dramatic reduction in the number of bits examined, which decreases monotonically with the threshold. Nonetheless, while reducing the number of bits examined, more severe model restrictions can also be expected to lead to an increased time complexity of fitness evaluation due to the larger population sizes required to achieve reliable performance.

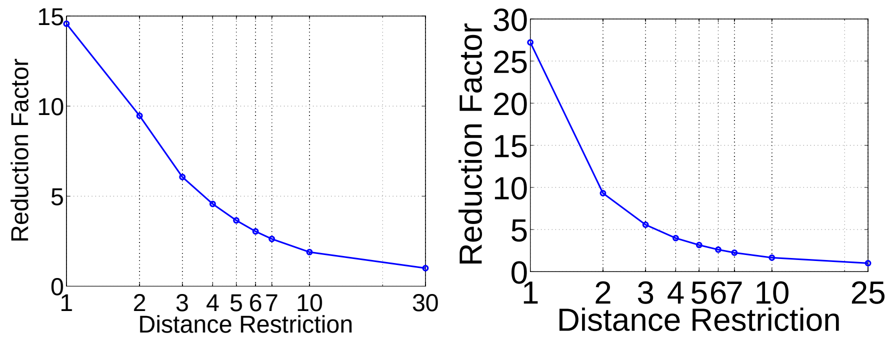
6.2 Results on rADPs

To examine distance-based model restrictions on rADPs, we considered two different combinations of values of n and o. For the heavier overlap of $o = 2$, we considered the problem of 92 bits. On the other hand, for overlap $o = 1$, we considered instances of $n = 101$ bits. In both cases, we set $k = 5$. 1000 random problem instances are tested for each problem size.



(a) $n = 92, k = 5, o = 2$ (b) $n = 101, k = 5, o = 1$

Figure 3: Execution speedup by distance restriction on rADPs.



(a) $n = 92, k = 5, o = 2$ (b) $n = 101, k = 5, o = 1$

Figure 4: Reduction in the number of bits examined by distance restriction on rADPs.

Figure 3a shows the execution time speedup for various distance thresholds on rADPs of $n = 92$ with $o = 2$. We see that the optimal speedup is obtained when we allow dependencies only between bits in the same subproblem or those that are part of adjacent, overlapping subproblems. On the other hand, figure 3b shows that for $n = 101$ and $o = 1$, the best speedup is obtained when we allow only dependencies between the bits in the same subproblem. Therefore, it is clear that stronger overlap leads to the need of considering dependencies between bits at greater distances; on the other hand, weaker overlap allows more severe model restrictions.

Figure 4 shows the reduction in the number of bits examined during the entire model-building procedure for both cases. We see that the results for both cases are very similar. Similarly as for the 3D spin glass, the number of bits examined decreases with stronger restrictions.

6.3 Results on Minimum Vertex Cover

To examine the distance-based model restrictions on MVC, we considered 1000 random instances of the $G(n, m)$ model with $n = 300$ nodes and $m = 4n$.

Figure 5 (left-hand side) shows the execution-time speedup for various distance thresholds on MVC. The speedups obtained are much smaller than those obtained for the other two test problems. Even in the best case (threshold of 4) we only see slightly more than a 10% improvement. While this is somewhat surprising, the reason can be explained by the obtained reduction of the number

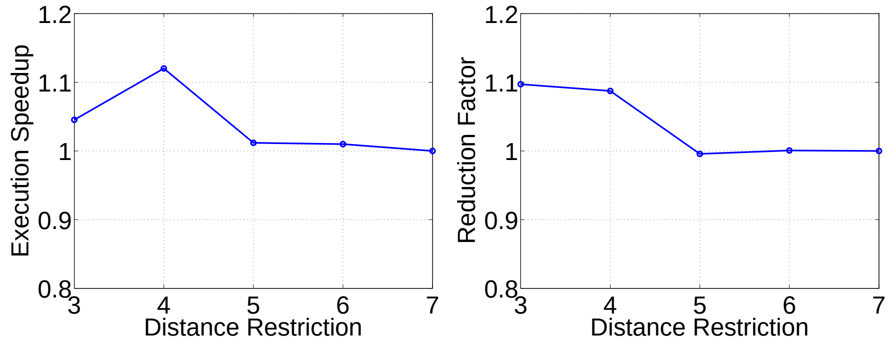


Figure 5: Execution time speedup and reduction in the number of bits examined for MVC.

dist	1	2	3	4	5	6	none
total	1200000	8388911	26212460	8883471	137371	580	151960
prob	2.7%	21.3%	79.6%	99.4%	99.7%	99.7%	100%

Table 1: Percentage of pairs of nodes with different distance thresholds for random instances of the $G(n, m)$ model with $n = 100$ and $m = 400$.

of bits examined, which is shown in Figure 5 (right-hand side). Unlike in the previous experiments, the reduction in the number of bits examined reaches only a factor of 1.1 even for the most severe distance restrictions. Furthermore, we see that there is a sharp drop in the number of bits examined between the thresholds of 4 and 5.

To examine the MVC results in more depth, we looked at the distribution of distances over all 1000 problem instances, which is shown in table 1. These results show that for thresholds of 5 or more, almost all pairs of nodes are allowed to be connected, which explains why the reduction in the number of bits examined is negligible. Furthermore, the results show that when decreasing the thresholds of 1 and 2 leads to a dramatic reduction in the number of possible dependencies, explaining poor performance of hBOA with such severe model restrictions.

7 Summary and Conclusions

This paper analyzed the speedups obtained in hBOA by defining a distance metric over problem variables and strictly disallowing dependencies between variables at a distance greater than a given threshold. Three test problems were considered: 3D spin glass, random additively decomposable problems (rADPs), and minimum vertex cover (MVC). It was shown that substantial speedups of 1.5 to 2.2 can be obtained for the 3D spin glass and rADPs, whereas only small speedups of about 1.1 can be obtained for MVC.

While we only looked at three problems, a distance metric can be defined for many other problems, including graph coloring, atomic cluster optimization, and the quadratic assignment problem. An important direction for future research is to extend the proposed techniques to other important classes of problems. Furthermore, since the benefits of using distance-based model restrictions depend on the distance threshold, an important direction for future work is to develop automated methods for choosing an appropriate threshold or to eliminate the threshold altogether using soft distance-based restrictions. Finally, similar ideas might be applied to other EDAs based on multivariate models.

An important thing to remember is that combining efficiency enhancement techniques can lead to multiplicative speedups (Goldberg, 2002; Sastry, Pelikan, & Goldberg, 2006). That means that even a moderate speedup of 1.5 or 2 can significantly contribute to the overall efficiency.

Acknowledgments

This project was sponsored by the National Science Foundation under CAREER grant ECS-0547013, by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant FA9550-06-1-0096, and by the University of Missouri in St. Louis through the High Performance Computing Collaboratory sponsored by Information Technology Services, and the Research Award and Research Board programs.

The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the Air Force Office of Scientific Research, or the U.S. Government. Some experiments were done using the hBOA software developed by Martin Pelikan and David E. Goldberg at the University of Illinois at Urbana-Champaign and most experiments were performed on the Beowulf cluster maintained by ITS at the University of Missouri in St. Louis.

References

Baluja, S. (1994). *Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning* (Tech. Rep. No. CMU-CS-94-163). Pittsburgh, PA: Carnegie Mellon University.

Baluja, S. (2006). Incorporating a priori knowledge in probabilistic-model based optimization. In Pelikan, M., Sastry, K., & Cantú-Paz, E. (Eds.), *Scalable optimization via probabilistic modeling: From algorithms to applications* (pp. 205–219). Springer.

Barahona, F. (1982). On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical, Nuclear and General*, *15*(10), 3241–3253.

Bollobas, B. (2001). *Random graphs*. Cambridge University Press.

Chickering, D. M., Heckerman, D., & Meek, C. (1997). *A Bayesian approach to learning Bayesian networks with local structure* (Technical Report MSR-TR-97-07). Redmond, WA: Microsoft Research.

Friedman, N., & Goldszmidt, M. (1999). Learning Bayesian networks with local structure. In Jordan, M. I. (Ed.), *Graphical models* (pp. 421–459). MIT Press.

Goldberg, D. E. (2002). *The design of innovation: Lessons from and for competent genetic algorithms*. Kluwer.

Harik, G. R. (1995). Finding multimodal solutions using restricted tournament selection. *International Conference on Genetic Algorithms (ICGA-95)*, 24–31.

Hauschild, M., Pelikan, M., Lima, C., & Sastry, K. (2007). Analyzing probabilistic models in hierarchical boa on traps and spin glasses. *Genetic and Evolutionary Computation Conference (GECCO-2007)*, *I*, 523–530.

Hauschild, M., Pelikan, M., Sastry, K., & Goldberg, D. E. (2008). Analyzing probabilistic models in hierarchical boa on traps and spin glasses. *Genetic and Evolutionary Computation Conference (GECCO-2008)*, *I*, 523–530.

Howard, R. A., & Matheson, J. E. (1981). Influence diagrams. In Howard, R. A., & Matheson, J. E. (Eds.), *Readings on the principles and applications of decision analysis*, Volume II (pp. 721–762). Menlo Park, CA: Strategic Decisions Group.

Karp, R. M. (1972). Reducibility among combinatorial problems. In Miller, R. E., & Thatcher, J. W. (Eds.), *Complexity of Computer Computations* (pp. 85–103). Plenum Press.

Larranaga, P., & Lozano, J. A. (Eds.) (2001). *Estimation of distribution algorithms: A new tool for evolutionary computation*. Springer.

Larrañaga, P., & Lozano, J. A. (Eds.) (2002). *Estimation of distribution algorithms: A new tool for evolutionary computation*. Boston, MA: Kluwer.

Lima, C. F., Pelikan, M., Sastry, K., Butz, M. V., Goldberg, D. E., & Lobo, F. G. (2006). Substructural neighborhoods for local search in the Bayesian optimization algorithm. *Parallel Problem Solving from Nature*, 232–241.

Mühlenbein, H., & Mahnig, T. (2002). Evolutionary optimization and the estimation of search distributions with applications to graph bipartitioning. *International Journal on Approximate Reasoning*, *31*(3), 157–192.

Mühlenbein, H., & Paaß, G. (1996). From recombination of genes to the estimation of distributions I. Binary parameters. *Parallel Problem Solving from Nature*, 178–187.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Pelikan, M. (2005). *Hierarchical Bayesian optimization algorithm: Toward a new generation of evolutionary algorithms*. Springer-Verlag.

Pelikan, M., & Goldberg, D. E. (2001). Escaping hierarchical traps with competent genetic algorithms. *Genetic and Evolutionary Computation Conf. (GECCO-2001)*, 511–518.

Pelikan, M., Goldberg, D. E., & Lobo, F. (2002). A survey of optimization by building and using probabilistic models. *Computational Optimization and Applications*, *21*(1), 5–20.

Pelikan, M., Kalapala, R., & Hartmann, A. K. (2007). Hybrid evolutionary algorithms on minimum vertex cover for random graphs. *Genetic and Evolutionary Computation Conference (GECCO-2007)*, 547–554.

Pelikan, M., & Sastry, K. (2004). Fitness inheritance in the Bayesian optimization algorithm. *Genetic and Evolutionary Computation Conf. (GECCO-2004)*, *2*, 48–59.

Pelikan, M., Sastry, K., Butz, M. V., & Goldberg, D. E. (2006a). *Hierarchical BOA on random decomposable problems* (MEDAL Report No. 2006001). Missouri Estimation of Distribution Algorithms Laboratory, Univ. of Missouri–St. Louis.

Pelikan, M., Sastry, K., Butz, M. V., & Goldberg, D. E. (2006b). Performance of evolutionary algorithms on random decomposable problems. *Parallel Problem Solving from Nature (PPSN IX)*, 788–797.

Pelikan, M., Sastry, K., & Goldberg, D. E. (2002). Scalability of the Bayesian optimization algorithm. *International Journal of Approximate Reasoning*, *31*(3), 221–258.

Pelikan, M., Sastry, K., & Goldberg, D. E. (2006). Sporadic model building for efficiency enhancement of hierarchical BOA. *Genetic and Evolutionary Computation Conf. (GECCO-2006)*, 405–412.

Santana, R., Larrañaga, P., & Lozano, J. A. (2007). The role of a priori information in the minimization of contact potentials by means of estimation of distribution algorithms. In Marchiori, E., Moore, J. H., & Rajapakse, J. C. (Eds.), *EvoBio*, Volume 4447 of *Lecture Notes in Computer Science* (pp. 247–257). Springer.

Sastry, K. (2001). *Evaluation-relaxation schemes for genetic and evolutionary algorithms*. Master's thesis, University of Illinois at Urbana-Champaign, Department of General Engineering, Urbana, IL.

Sastry, K., Goldberg, D. E., & Pelikan, M. (2001). Don't evaluate, inherit. *Genetic and Evolutionary Computation Conf. (GECCO-2001)*, 551–558.

Sastry, K., Pelikan, M., & Goldberg, D. E. (2006). Efficiency enhancement of estimation of distribution algorithms. In Pelikan, M., Sastry, K., & Cantú-Paz, E. (Eds.), *Scalable Optimization via Probabilistic Modeling: From Algorithms to Applications* (pp. 161–185). Springer.

Sastry, K., Pelikan, M., & Goldberg, D. E. (2007). Empirical analysis of ideal recombination on random decomposable problems. *Genetic and Evolutionary Computation Conference (GECCO-2007)*, 1388–1395.

Schwarz, J., & Ocenasek, J. (2000). A problem-knowledge based evolutionary algorithm KBOA for hypergraph partitioning. Personal communication.

Weigt, M., & Hartmann, A. K. (2001). Minimal vertex covers on finite-connectivity random graphs - a hard-sphere lattice-gas picture. *Physical Review E*, *63*, 056127.