Estimation of Distribution Algorithm Based on Archimedean Copulas

Li-Fang Wang[1,2]
wlf1001@163.com

Jian-Chao Zeng[2]
zengjianchao@263.net

Yi Hong[1]
yudongmei@cnn.cn

1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, 730050, China

2. Complex System and Computational Intelligence Laboratory, Taiyuan University of Science and Technology, Taiyuan, 030024, China

ABSTRACT
Both Estimation of Distribution Algorithms (EDAs) and Copula Theory are hot topics in different research domains. The key of EDAs is modeling and sampling the probability distribution function which need much time in the available algorithms. Moreover, the modeled probability distribution function can not reflect the correct relationship between variables of the optimization target. Copula Theory provides a correlation between univariable marginal distribution functions and the joint probability distribution function. Therefore, Copula Theory could be used in EDAs. Because Archimedean copulas possess many nice properties, an EDA based on Archimedean copulas is presented in this paper. The experimental results show the effectiveness of the proposed algorithm.

Categories and Subject Descriptors
G.1.6 [**Numerical Analysis**]: Optimization—global optimization;

G.3 [**Probability and statistics**]: Random number generation ;

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms
Algorithms, Theory.

Keywords
Estimation of distribution algorithms (EDAs), Copula Theory, Archimedean copulas, Sklar's theorem

1. INTRODUCTION
Estimation of Distribution Algorithms (EDAs) catches many researchers' eyes since it was proposed [1] because of its new search strategy and its power to solve hard optimization problems. EDAs deriving from Genetic Algorithm are different from it. There are not cross operator and mutation operator in EDAs, which are replaced by modeling and sampling the probability distribution of the selected population. Thus, modeling and sampling the probability distribution is the kernel of EDAs. The classical algorithms are PBIL [2,3] , UMDA [4,5] , MIMIC [6,7], BOA [8], ENGA [9], etc. EDAs has been applied in many research areas such as multiobjective optimization [10,11,12], Flow Shop [13] and so on [14,15].

GEC'09, June 12-14, 2009, Shanghai, China.
ACM 978-1-60558-326-6/09/06.

There are two problems in the available EDAs. One is that the probability function of the promising population could not be estimated correctly, especially for the multivariate-dependant optimization problem; the other is that the process of modeling the probability distribution function is an optimization problem and need much time to optimize.

Copula Theory is popular in statistical area [16,17] and financial area [18,19] because it provides a way of studying scale-free measures of dependence. But the research of copula in computation intelligence is not available until now. Copulas join multivariate distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval (0,1)[16]. An EDA modeling and sampling the probability distribution in the light of Copula Theory will save the operation time and reflect exactly the dependence of variables.

In section 2, a brief description of copula theory is provided. In section 3, the EDAs based on Archimedean copulas (Archimedean Copula EDAs) is discussed. The results of experiments are presented in section 4 and the conclusions are provided in section 5.

2. COPULA THEORY
The definition of copula and an essential theorem in copula theory are provided in the following.

Definition 1: Let $\mathbf{I}$=[0,1]. A two-dimensional copula (or 2-copula, or briefly, a copula) is a function C from $\mathbf{I}^2$ to $\mathbf{I}$ with the following properties:

1) For every u, v in I,
$$C(u,0)=0=C(0,v) \tag{1}$$
and
$$C(u,1)= u \text{ and } C(1,v) =v; \tag{2}$$

2) For every u_1,u_2,v_1,v_2 in $\mathbf{I}$ such that $u_1 \le u_2$ and $v_1 \le v_2$,
$$C(u2, v2)-C(u2, v1)-C(u1, v2)+C(u1, v1)\ge 0. \tag{3}$$

Sklar's theorem: Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x,y in $\mathbf{R}$,
$$H(x,y) = C(F(x),G(y)). \tag{4}$$
If F and G are continuous, then C is unique; otherwise, C is uniquely determined on Ran$F$×Ran$G$. Conversely, if C is a copula and F and G are distribution functions, then the function H defined by (4) is a joint distribution function with margins F and G.

Sklar's theorem plays an important role in Copula Theory, and the argument could be found in [16]. Copulas are divided into two classes, elliptical copulas and Archimedean copulas. Archimedean copulas find a wide range of applications because of the ease with which they can be constructed and the many nice properties possessed by the members of this class. [16] lists a great variety of Archimedean copulas.

3. ARCHIMEDEAN COPULA EDAs

According to Sklar's theorem, two steps are performed in order to construct the joint probability distribution function of a random vector. The first step is constructing the margins of each random variable separately. The second step is selecting a proper copula to construct the joint distribution. Therefore, the distribution character of each random variable and their relationship can be studied by themselves. This way can be used in EDAs to model the joint probability distribution function. And then samples are generated from the specified joint distribution by use of the copula.

The optimization problem is

$$\min f(X) = f(x_1, x_2)\ ,\ \ x_i \in [a_i, b_i] \quad (i = 1,2)\,. \tag{5}$$

Denote the selected population with size s as

$$\mathbf{x} \triangleq \{x^i = (x^i_1, x^i_2), i = 1, 2, \ldots, s\} \tag{6}$$

In other words, $\mathbf{x}$ are the s observations of the random vector (X_1, X_2). The marginal distribution function of each random variable X_i can be estimated by normal distribution, t-distribution or empirical distribution, etc. Denote the marginal distribution function of X_i as $u=F(x_1)$ and $v=G(x_2)$. The joint probability distribution function is constructed with a selected copula C and the estimated margins in the light of Sklar's theorem.

The next step is generating samples from the joint distribution using the copula as a tool. By virtue of Sklar's theorem, it is need only to generate a pair (u,v) of observations of uniform (0,1) random variables (U,V) whose joint distribution function is C, and then transform those uniform variates via the quasi-inverse of the marginal distribution functions. One procedure for generating such of a pair (u,v) of uniform (0,1) variates is the conditional distribution method. For this method, the conditional distribution function for V given $U = u$ is need, which is denoted as $C_u(v)$:

$$C_u(v) = P(V \le v \mid U = u) = \lim_{\Delta u \to 0} \frac{C(u+\Delta u, v) - C(u,v)}{\Delta u} = \frac{\partial C(u,v)}{\partial u}\,. \tag{7}$$

$C_u(v)$ exists and is non-decreasing almost everywhere in **I**.

Conclusively, the generation of sample is performed as the following steps:

s1. Generate two independent uniform (0,1) variates u and t;

s2. Set $v = C_u^{(-1)}(t)$, where $C_u^{(-1)}(t)$ denotes a quasi-inverse of $C_u(v)$.

s3. The desired pair is (u,v).

s4. Set $x_1=F^{(-1)}(u)$, $x_2=G^{(-1)}(v)$, then (x_1,x_2) is a sample of the specified joint distribution.

To sum up, the process for implementing 2-D Copula-EDA is as follows:

s1: Initialize (***pop***, *N*). Randomly generate initial population pop with size *N*. set generation count $g \leftarrow 0$.

s2: Selection (***pop***, ***spop***, *select-rate*). Select the best *select-rate*×*N* agents form pop to ***spop*** according to the agents' fitness.

s3: Copula-generator (***pop***, ***spop***, *mutate-rate*).

s3.1: Construct the distribution model of ***spop***;

s3.2: Generate a new population based on the specified joint distribution, and randomly generate some agents by the rate *mutate-rate*.

s4: Stop if the termination criterion is met.

s5: Set $g \leftarrow g+1$, and then go to s2.

4. EXPERIMENTS

The following functions are used to test the effectiveness of the proposed algorithm. $f_1 \sim f_3$ and $f_6 \sim f_8$ are adopted from [3].

- $\min f_1 = -\dfrac{100}{10^{-5} + \sum_i |y_i|}$, $y_1 = x_1, y_i = x_i + y_{i-1} (i \ge 2)$, $x_i \in [-3,3]$, the optimal is $f_1{*}(0,0,\ldots0) = -10^7$.
- $\min f_2 = -\dfrac{100}{10^{-5} + \sum_i |y_i|}$, $y_1 = x_1, y_i = x_i + \sin y_{i-1} (i \ge 2)$, $x_i \in [-3,3]$, the optimal is $f_2{*}(0,0,\ldots0) = -10^7$.
- $\min f_3 = -\dfrac{100}{10^{-5} + \sum_i |y_i|}$, $y_i = 0.024 \times (i+1) - x_i$, $x_i \in [-3,3]$, the optimal is $f_3{*}(0.024 \times 2, 0.024 \times 3, \ldots 0.024 \times (n+1)) = -10^7$.
- $\min f_4 = \sum_i x_i^2$, $x_i \in [-500,500]$, the optimal is $f_4{*}(0,0,\ldots0) = 0$.
- $\min f_5 = 1 + \sum_i (\sin x_i)^2 - 0.1\exp(-\sum_i x_i^2)$, $x_i \in [-10,10]$, the optimal is $f_5{*}(0,0,\ldots0) = 0.9$.
- $\min f_6 = \sum_i (x_i^2 - A\cos(2\pi x_i) + A)$, $x_i \in [-5,5]$, the optimal is $f_6{*}(0,0,\ldots0) = 0$.
- $\min f_7 = \sum_i (418.9829 + x_i \sin\sqrt{|x_i|})$, $x_i \in [-500,500]$, the optimal is $f_7{*}(-420.9687, -420.9687, \ldots -420.9687) = 0$.
- $\min f_8 = \sum_i x_i^2 - \prod_i \cos(\frac{x_i}{\sqrt{i+1}})$, $x_i \in [-100,100]$, the optimal is $f_8{*}(0,0,\ldots0) = -1$.
- $\min f_9 = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$, $x_1, x_2 \in [-2,2]$, the optimal is $f_9{*}(0,-1) = 3$.

The following two Archimedean copulas are chosen.

- $C_1(u,v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \theta \ge -1, \theta \ne 0$
- $C_2(u,v) = \dfrac{uv}{1 - \theta(1-u)(1-v)}, -1 \le \theta < 1$

All the one-dimensional marginal distributions are normal distributions. Table 1 displays the experimental results.

Table 1. Experimental results of Archimedean Copula EDAs

Test Function	Copula	Convergence Rate	Convergence Generation
f_1	C_1	50/50	13.7400
	C_2	50/50	13.0600
f_2	C_1	50/50	13.6800
	C_2	50/50	13.1400
f_3	C_1	50/50	4.8000
	C_2	50/50	11.7200
f_4	C_1	50/50	16.2800
	C_2	50/50	16.7000
f_5	C_1	50/50	75.7800
	C_2	50/50	76.7800
f_6	C_1	50/50	43.6800
	C_2	50/50	28.7400
f_7	C_1	47/50	55.5957
	C_2	50/50	32.8600
f_8	C_1	50/50	14.9600
	C_2	50/50	15.1200
f_9	C_1	48/50	12.2083
	C_2	50/50	12.2600

All test functions are optimized in 2-dimensional spaces, the maximal generation g is set to 1000. The search terminates if the distance between the best solution found so far and the optimum is less than the predefined precision (10^{-5} for other test functions in spite of 10^{-3} for f_7). Parameters are set to (*select-rate*=0.2, *mutate-rate*=0.05, population size N=100) for all experiments. The convergence rate and the convergence generations are the average results of 50 runs. The experimental results show that Copula-EDA converges to the global optimum quickly in the test functions. There is not much difference in performance between two copulas for other test functions despite f_3 and f_6. Both the algorithms proposed in this paper perform better than the copula-EDA based on Gaussian copula and PBILc [20].

5. CONCLUSION

Compared with GAs, EDAs utilizes well the information provided from the promising population, and becomes the hot topic of Intelligence Computation. Whereas, it is a complex process to model and sample the probability distribution of the promising population. Copula theory in statistics provides an easier way for it. The process of modeling the probability distribution can be divided into modeling univariate margins and selecting a copula. Sampling from the constructed model can also be done by use of copula. From the experimental results it is obvious that the Archimedean Copula EDAs proposed in this paper is effective. But two-dimensional Archimedean copula EDAs is only an attempt to join EDAs with copulas. The multi-dimensional algorithm is the next target of our study.

6. ACKNOWLEDGMENTS

This work was supported in part by the Youth Research Fund of Taiyuan University of Science and Technology (No.2007116), the Chinese Nature Science Fund (No. 60674104) and the Youth Research Fund of ShanXi province (No. 2006021019).

7. REFERENCES

[1] Larranaga, P., and Lozano, J. A. 2002. Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation. Boston: Kluwer Academic Publishers.

[2] Baluja, S. 1994. Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning. Technical Report CMU-CS-94-163, Pittsburgh, PA: Carnegie Mellon University. URL= http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.8084.

[3] Sebag, M., and Ducoulombier, A. 1998. Extending population-based incremental learning to continuous search spaces. In Proceedings of the 5th Conference on Parallel Problem Solving from Narute- PPSN V (Amsterdam, Netherlands, September 27-30, 1998). Volume 1498 of Lecture Notes in Computer Science. Springer, Berlin, 418-427. URL= http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.1884.

[4] Muhlenbein, H., and Paass, G. 1996. From recombination of genes to the estimation of distributions I. Binary Parameters. In Proceedings of the 4th Conference on Parallel Problem Solving from Nature- PPSN IV (Berlin, Germany, September 22 - 26, 1996). Volume 1141 of Lecture Notes in Computer Science. Springer, Berlin, 178-187. URL= http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.7.7030&rep=rep1&type=ps.

[5] Larranaga, P., Etxeberria, R., Lozano, J. A., and Pena, J. M. 2000. Optimization in continuous domains by learning and simulation of Gaussian networks. In Proceedings of the Genetic and Evolutionary Computation Conference (Las Vegas, Nevada, USA, July 8-12, 2000). GECCO '00 .Morgan Kaufmann, San Francisco, 201-204.

[6] De Bonet, J. S., Isbell, C. L. , and Viola, P. 1996. MIMIC: Finding optima by estimation probability densities. Advances in Neural Information Processing Systems , Cambridge: MIT Press, 9:424-430. URL= http://books.nips.cc/papers/files/nips09/0424.pdf

[7] Zhong, W., Liu, J., Liu, F., and Jiao, L. 2004. Second order estimation of distribution algorithms based on kalman filter, Chinese J. Comput., September 2004, 27(9):1272-1277 (in Chinese)

[8] Pelikan, M., Goldberg, D. E., and Cantu-Paz, E. 1999. BOA: the Bayesian optimization algorithm. In Proceedings of the Genetic and Evolutionary Computation Conference (Orlando, Florida, USA, July 13-17, 1999,). GECCO 1999, .Morgan Kaufmann, San Francisco, 525-532. URL=

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.2148&rep=rep1&type=pdf

[9] Larranga, P., Etxeberria, R., Lozano, J. A., and Pena, J. M. 1999. Optimization by Learning and Simulation of Bayesian and Gaussian Networks. Technical Report EHU-KZAA-IK-4/99, University of the Basque Country, Spain. URL= http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.1895

[10] Bosman P.A.N. and Thierens D. 2006. Multi-objective optimization with the naive MIDEA. In Towards a New Evolutionary Computation. Advances in Estimation of Distribution Algorithms, J.A. Lozano, P. Larrañaga, I. Inza and E. Bengoetxea, Ed. Springer-Verlag, Berlin, 123-157.URL= http://www.springerlink.com/content/j9n0ul146357r552/

[11] Nazan, K., Goldberg, D. E., and Pelikan, M. 2002. Multi-objective Bayesian optimization algorithm. IlliGAL Report No.2002009, University of Illinois at Urbana-Champaign, Urbana, Illinois,

[12] Pelikan, M., Sastry, K., and Goldberg, D. E. 2005. Multiobjective hBOA, clustering, and scalability. In Proceedings of the Genetic and Evolutionary Computation Conference (Washington DC, USA, June 25-26, 2005) GECCO'05. ACM Press, New York, 663-670. URL= http://portal.acm.org/citation.cfm?id=1068122.

[13] Salhi, A., Rodriguez, J. A. V., and Zhang, Q. 2007. An Estimation of Distribution Algorithm with Guided Mutation for a Complex Flow Shop Scheduling Problem. In Proceedings of the Genetic and Evolutionary Computation Conference (London, England, UK, July 7-11, 2007). ACM Press, New York, 570 – 576. URL= http://portal.acm.org/citation.cfm?id=1276958.1277076

[14] Simionescu, P. A., Beale, D. G., and Dozier, G. V. 2006. Teeth-number synthesis of a multispeed planetary transmission using an estimation of distribution algorithm. J. Mech. Design., January 2006, 128(1):108-115.

[15] Santarelli, S., Yu, T., Goldberg, D. E., Altshuler, E., O'Donnell, T., and Southall H. 2006. Military antenna design using simple and competent genetic algorithms. Math. Comput. Model., 43(9-10):990-1022

[16] Nelsen, R. B. 2006. An Introduction to Copulas (Second Edition). Springer, New York.

[17] Demarta, S., and McNeil, A. J. 2007. The t Copula and Related Copulas, Int. Stat. Rev.**,**73(1):111-129.

[18] De Giuli, M. E., Fantazzini, D., and Maggi, M. A. 2008. A New Approach for Firm Value and Default Probability Estimation beyond Merton Models, Comput. Econ., March 2008, 31(2):161–180

[19] Cherubini, U., Luciano, E., and Vecchiato, W. 2004. Copula methods in finance. John Wiley.

[20] Wang, L. F., Zeng, J. C., and Hong, Y. Estimation of Distribution Based on Copula Theory. In the 2009 IEEE Congress on Evolutionary Computation (Trondheim, Norway, May 18-21, 2009). Paper 622. In press.