The Directional EDA for Global Optimization

Pedro Mayorga Alvarez
Centre for Research in Mathematics,
Department of Computer Science
AP. 402, Guanajuato, Gto.
CP. 36000, México
mayorga@cimat.mx

Arturo Hernández Aguirre
Centre for Research in Mathematics,
Department of Computer Science
AP. 402, Guanajuato, Gto.
CP. 36000, México
artha@cimat.mx

ABSTRACT

This article presents a robust EDA for global optimization with real parameters. The approach is based on the linear combination of individuals of two populations. One is the current population P_t, from which a probability density model is created and a new population P_s is simulated. The new population P_{t+1} is a linear combination of P_t and P_s. The linear combination factor involved is self-adaptive.

Categories and Subject Descriptors: J.2[Physical Sciences and Engineering]Mathematics and Statistics

General Terms: Algorithms, Design, Performance.

Keywords: Estimation Distribution Algorithms, Convergence, Optimization.

1. INTRODUCTION

Premature convergence is a well known issue of Estimation Distribution Algorithms (EDAs) [1]. The approach of this paper is to keep diversity by combining two populations. The first population $\mathcal{P}^{(t)}$ is the one available at the current generation. A sample of the bests is taken and a probability density function (PDF) model is created. The second population is simulated from the model, and then linearly combined with the first one, individual by individual, resulting in a new individual that populates the new generation $\mathcal{P}^{(t+1)}$. This paper presents the Directional EDA (DEDA).

2. DIRECTIONAL EDA

Consider a vector $u \in \mathcal{P}^{(t)}$ that will get an increment and change its position to $v \in \mathcal{P}^{(t+1)}$. The step size h is: $h = v - u$. In a more general situation, we can take a step λh resulting in the new position $v^* = u + \lambda h$. In general $v^* = u + \lambda(v - u)$. The new vector $\lambda(v - u)$ is the *direction* of motion biased by the best individuals seeking the optimum. Algorithm 1 presents the pseudocode of the Directional EDA (named after the direction vector). The adaptation of the linear coefficient takes place in steps 13 through 17. Notice that each individual has its own coefficient. Every coefficient is adjusted when all three past function values were either incremented (step 13) or decremented (step 15).

Algorithm 1 Pseudocode of the proposed Directional EDA

1: Define the set of three elements with last values of function $f(\cdot)$ of the ith individual of the population at time t: $\mathcal{F}_i^{(t)} = \{f_i^k | k = t, t-1, t-2\}$, where $f_i^k \equiv f(\mathcal{P}_i^{(k)})$
2: $t \leftarrow 1$ and $\Lambda_i^{(1)} \leftarrow 1, i = 1 \dots n$.
3: Set $\mathcal{P}^{(1)}$ as random population of size n.
4: **repeat**
5: Sorting $\mathcal{P}^{(t)}$ in ascending order, respect to $f(\cdot)$.
6: Set $\mathcal{Q}^{(t)}$ of size m, as the subset of the first elements of $\mathcal{P}^{(t)}$.
7: Estimate: PDF $\mathcal{H}^{(\mathbf{t})}$ such that $\mathcal{Q}^{(t)} \sim \mathcal{H}^{(\mathbf{t})}$.
8: Generate: new population of size n from the model $\mathcal{H}^{(\mathbf{t})}$: $\mathcal{S}^{(t)} \sim \mathcal{H}^{(\mathbf{t})}$.
9: **for** $i = 1 \dots n$ **do**
10: $u \leftarrow \mathcal{P}_i^{(t)}$, $v \leftarrow \mathcal{S}_i^{(t)}$, $\lambda \leftarrow \Lambda_i^{(t)}$
11: Calculate: $v^* = u + \lambda(v - u)$.
12: $\mathcal{P}_i^{(t+1)} \leftarrow v^*$
13: **if** The set $\mathcal{F}_i^{(t)}$ is in ascending order **then**
14: $\Lambda_i^{(t+1)} \leftarrow 2\lambda$ {Increase Acceleration of step length}
15: **else if** The set $\mathcal{F}_i^{(t)}$ is in descending order **then**
16: $\Lambda_i^{(t+1)} \leftarrow \frac{1}{2}\lambda$ {Decrease Acceleration of step length}
17: **else**
18: $\Lambda_i^{(t+1)} \leftarrow \Lambda_i^{(t)}$
19: **end if**
20: **end for**
21: $t \leftarrow t + 1$
22: **until** Termination

3. EXPERIMENTS AND RESULTS

A PDF model is approximated using a Gaussian Mixture Model (GMM) with four kernels. The GMM is calculated with the Expectation Maximization (EM) algorithm and it is initialized with the K-means algorithm. Each problem was ran 10 times. Population size=20, and sample size=15.

3.1 Experiment 1

The results for a well known 5-functions benchmark are shown in Table 1. The stop criterion was set to reach an error smaller than $1.0E-6$.

3.2 Experiment 2

The goal is to solve and compare results for seven multimodal functions listed in [2]. The detailed settings are listed in Table 2. The stop criteria was set to reach 301850 evaluations or when the result obtained was closer than a $\delta =$

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-130-9/08/07.

Problem	Best Approximation	Evaluations
Dimension 10		
Sum-Can	1.0000E+5 ± 1.5789E-7	3864 ± 248
Griewangk	6.6407E-7 ± 2.9451E-7	1060 ± 268
Sphere	6.5735E-7 ± 2.4091E-7	1056 ± 193
Rosenbrock	5.0950E-7 ± 3.8759E-7	20264 ± 15675
Ackley	7.9657E-7 ± 1.4341E-7	1616 ± 109
Dimension 50		
Sum-Can	1.0000E+5 ± 1.1506E-7	4388 ± 363
Griewangk	6.0690E-7 ± 2.4766E-7	964 ± 175
Sphere	6.8914E-7 ± 1.9606E-7	1188 ± 149
Rosenbrock	4.8490E-7 ± 3.6665E-7	47400 ± 54566
Ackley	8.5737E-7 ± 1.0040E-7	1588 ± 121

Table 1: Best individual and number of function evaluations for Experiment 1 after 10 runs

Function	D	Domain	Type	Optimum
Sphere	30	$[-100, 100]$	Min	0
Sum-Can	10	$[-0.16, 0.16]$	Max	10^5
TwoPeaks	5	$[-100, 100]$	Max	10.1053
ThreePeaks	5	$[-100, 100]$	Max	10.1053
Shekel ($n = 5$)	4	$[0, 10]$	Max	10.1033
Shekel ($n = 5$)	30	$[0, 10]$	Max	10.0139
Schwefel	30	$[-500, 500]$	Min	-12569.4866

Table 2: Test functions of Experiment 2

$[1.0e-7, 1.0e-7, 1.0e-4, 1.0e-4, 1.0e-4, 1.0e-4, 1.0e-2]$ (in the same order top to bottom of problems in Table 2). The experimental results are summarized in Table 3 (MFE stands for mean number of fitness function evaluations).

Problem	MFE	Best	Mean	S.D.
Sphere	2448	5.1633E-8	6.9286E-8	1.634E-8
Sum-Can	8929	1.0000E+5	1.0000E+5	1.452E-8
TwoPeaks	24189	10.1053E00	10.1053E00	4.456E-5
ThreePeaks	33151	10.1053E00	10.1053E00	4.371E-5
Shekel (D=4)	7748	10.1033E00	10.1032E00	2.952E-5
Shekel (D=30)	53024	10.0139E00	9.8643E00	4.728E-1
Schwefel	240000	-12569.48	-1.0852E+4	1.819E+3

Table 3: DEDA results for Experiment 2.

3.3 Experiment 3.

The functions of this experiment are convex and monotone [1]. The goal of the experiment is to estimate the scalability of the algorithm through the linear regression coefficient. The average number of evaluations is measured versus dimensionality values 2,4,8,10,20,40 and 80. The functions are defined in Table 4.

4. COMMENTS AND CONCLUSIONS

Notice that the non linear Rosenbrock function is perfectly solved in dimensions 10 and 50 in Experiment 1. Also the function Sum-Can is solved in only 3864 and 4388 fitness function evaluations. In Experiment 2, DEDA easily solved the problems that required automated clustering techniques in [2]. For Experiment 3, the plots in Figure 1 denoting scalability are almost flat. The comparison in Table 5 shows the regression coefficients are really superior for the proposed directional EDA than for the adaptive variance IDEA [1].

Name	**Definition**	**Value to reach**
Sphere	$\sum_{i=1}^{l} x_i^2$	10^{-10}
Ellipsoid	$\sum_{i=1}^{l} 10^{6\frac{i-1}{l-1}} x_i^2$	10^{-10}
Cigar	$x_1^2 + \sum_{i=2}^{l} 10^6 x_i^2$	10^{-10}
Tablet	$10^6 x_1^2 + \sum_{i=2}^{l} x_i^2$	10^{-10}
Cigar Tablet	$x_1^2 + \sum_{i=2}^{l-1} 10^4 x_i^2 + 10^8 x_l^2$	10^{-10}
Two Axes	$\sum_{i=1}^{\lfloor l/2 \rfloor} 10^6 x_i^2 + \sum_{i=\lfloor l/2 \rfloor}^{l} x_i^2$	10^{-10}
Different Powers	$\sum_{i=1}^{l} \lvert x_i^2 \rvert^{2+10\frac{i-1}{l-1}}$	10^{-15}
Rosenbrock	$\sum_{i=1}^{l-1} \left(100\left(x_i^2 - x_{i+1}\right)^2 + (x_i - 1)^2\right)$	10^{-10}
Parabolic Ridge	$-x_1 + 100 \sum_{i=2}^{l} x_i^2$	10^{-10}
Sharp Ridge	$-x_1 + 100\sqrt{\sum_{i=2}^{l} x_i^2}$	10^{-10}

Table 4: Functions and max error for Experiment 3

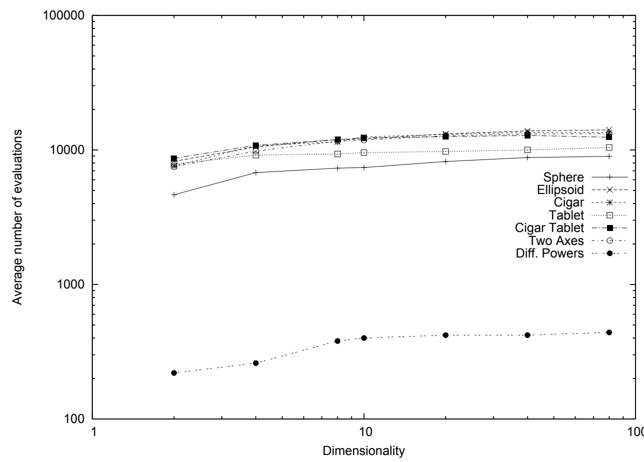

Figure 1: Dimensionality versus average number of fitness function evaluations

Function	Best result in [1] Algorithm	β	This Paper β
Sphere	CMA-ES	0.9601	0.1572
Ellipsoid	IDEA	1.2171	0.1359
Cigar	CMA-ES	1.1093	0.1374
Tablet	IDEA	1.0806	0.0647
Cigar Tablet	IDEA	1.1142	0.0875
Two Axes	IDEA	1.2854	0.1421
Different Powers	AVS-IDEA	1.1692	0.1861
Parabolic Ridge	CMA-ES	1.0853	0.0995

Table 5: Results for Experiment 3, $\log e = \epsilon + \log l^{\beta}$.

5. REFERENCES

[1] J. Grahl, P. A. N. Bosman, and F. Rothlauf. The correlation-triggered adaptive variance scaling IDEA. In *GECCO '06: Proceedings of the 8th annual conference on Genetic and Evolutionary Computation*, pages 397–404. ACM Press, 2006.

[2] Q. Lu and X. Yao. Clustering and learning gaussian distribution for continuos optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 35(2):195–204, May 2005.