Evolutionary Algorithms with Mutations based on α-Stable and Directional Distributions

Andrzej Obuchowicz[1]

[1] University of Zielona Góra, Institute of Control and Computation Engineering, Zielona Góra, Poland, e-mail: A.Obuchowicz@issi.uz.zgora.pl

Abstract. In this paper, a concept of directional mutations for phenotypic evolutionary algorithms is presented. The proposed approach allows, in a very convenient way, to adapt the probability measure underlying the mutation operator during evolutionary process. Simulated experiments confirms the thesis that proposed mutation improves the effectiveness of evolutionary algorithms in the case of the local as well as global optimization problems.

1 Introduction

Most applications of Evolutionary Algorithms (EAs) which use the floating point representation of population individuals use the Gaussian mutation as a mutation operator (cf. [1, 2]). A new individual is obtained by adding a normally distributed random value to each entry of a selected parent. The choice is usually justified by the Central Limit Theorem. Mutations in nature are caused by a variety of physical and chemical factors that are not identifiable or measurable. These factors are considered as independent and identically distributed (i.i.d.) random perturbations. The Generalized Central Limit Theorem states that the only possible nontrivial limit of normalized sums of i.i.d. terms is Lévy-stable, called also α-stable or just stable in the mathematical literature [3, 7, 14]. If the Lindeberg condition is obeyed, i.e., the first two absolute moments are finite, then the α-stable distribution reduces to the Gaussian distribution.

The suggestion that the application of α-stable distributions other than the Gaussian and the Cauchy distribution can be very attractive for evolutionary algorithms with the floating-point representation of individuals was first introduced by Gutowski [4]. Such an application of α-stable distributions to EAs based on the floating point representation of individuals has been simultaneously presented and studied by two independent groups of researches [5, 9]. Lee and Yao [5] apply the evolutionary programming algorithm with non-isotropic α-stable mutations to a set of 14 global optimization problems. Obuchowicz and Prętki [9] study the influence of some poorly known properties, called the *symmetry* and *surrounding effects*, of non-isotropic α-stable mutations on the exploitation and exploration abilities of EAs. In order to avoid negative influence of the above effects the isotropic α-stable mutation is proposed [11].

Spherically symmetric distributions guarantee that there is no preferable direction in the search space, which is a desired property especially at the beginning of the optimization process. Moreover, this also means that the effectiveness of an optimization technique does not depend on a reference frame. This, in the evolutionary algorithms (EAs) framework, was intensively studied by Obuchowicz [8]. An isotropic sampling strategy is effective for the low-dimensional problems, but its efficiency drastically decreases for large-scale problems [12]. Thus, many evolutionary algorithms are supplied with auxiliary heuristics which try to neutralize a negative influence of the dimensionality problem [8, 13]. An intuitive idea of dealing with the problem is to adjust a probability measure on the basis of information gained during the optimization process in order to calculate the preferable directions of mutations [8]. Such a special class of the so-called directional distributions is introduced in [12]. This new class gives an access to constructing new techniques for adaptation probability measures in the mutation operator.

In this paper the phenotype evolution with mutations based on α-stable and directional distributions proposed in [12] is considered. Simulation experiments are carried out in order to analyze the influence of algorithm exogenous parameters on its effectiveness of the global optimum searching for a set of benchmark optimization problems.

The paper is organized as follows. In the next section the α-stable distributions and isotropic multidimensional α-stable mutation are described as well as the concept of directional distributions with rotational symmetry is introduced. In the third section the description of a set of simulated experiments as well as compressed notes of observations and remarks are presented. Finally, the last section concludes the paper.

2 Lévy-stable distributions

2.1 Characteristic function representation

The ch.f. (characteristic function) of the α-stable distribution is parameterized by a quadruple $(\alpha, \beta, \sigma, \mu)$ [14], where α $(0 < \alpha \leq 2)$ is a stability index (tail index, tail exponent or characteristic exponent), β $(-1 \leq \beta \leq 1)$ is a skewness parameter, σ $(\sigma > 0)$ is a scale parameter and μ is a location parameter. The lack of closed-form formulas for probability density functions (pdfs) for all but three LSDs (Gaussian, Cauchy and Lévy distributions) has been a major drawback in the use of α-stable distributions by practitioners. Fortunately, there exist algorithmic formulas for simulating α-stable variables as well as computer programs to compute α-stable densities, distribution functions and quantiles [7].

If $\beta = 0$, then non-skewed α-stable distributions, called the Symmetric α-Stable ($S\alpha S$) distribution, is obtained. Thus, $Z \sim S_\alpha(0, \sigma, \mu) = S\alpha S(\sigma, \mu)$ (symmetric α-stable) can be expressed by

$$Z = \mu + \sigma X, \tag{1}$$

where $X \sim S\alpha S(1, 0) = S\alpha S$ has standardized symmetric α-stable distribution. The ch.f. of X is given by

$$\phi(k) = \exp\left(-|k|^\alpha\right). \tag{2}$$

For $\alpha = 1$, it is a ch.f. of the Cauchy distribution $C(0, 1)$, and for $\alpha = 2$, it becomes the ch.f. of the normal distribution $N(0, 1)$.

2.2 Multivariate isotropic α-stable distribution

The class of spherically symmetric distributions can be defined in a number of equivalent ways [3]. The most applicable one seems to be the following. The random variable

$\boldsymbol{X}$ can be generated as a result of the decomposition

$$\boldsymbol{X} = R\boldsymbol{U}^{(n)}, \tag{3}$$

where R and $\boldsymbol{U}^{(n)}$ are called, respectively, the generating variate and the uniform base of spherical distribution. If the generating variate R has $S\alpha S$ distribution, one says that the random vector $\boldsymbol{X} \sim \mathbf{S}\alpha\mathbf{SU}$ has *isotropic multivariate symmetry α-stable distribution.* Probably the simplest way to obtain the random vector uniformly distributed on the surface of a unit sphere $\mathbf{U}^{(n)}$ is described by following formula:

$$\boldsymbol{U}^{(n)} = \frac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}, \tag{4}$$

where $\boldsymbol{Y}$ is an n-dimensional, normally distributed random vector $\mathcal{N}_n(\mathbf{0}, \boldsymbol{I}_n)$ (where $\boldsymbol{I}_n$ stands for the identity covariance matrix of the order n).

2.3 Directional distribution

It can be proved [12] that the eficiency of an isotropic sampling strategy drastically decreases for large-scale problems. This problem can be neutralized with help of a technique which is based on the so-called directional distributions. In the literature, several classes of such distributions can be found [6], while for the need of evolutionary computations, the class of the so-called rotationally symmetric distributions $\mathcal{M}$ seems to be very attractive. The class $\mathcal{M}$ is usually parameterized by a pair $\{\boldsymbol{\mu}, \kappa\}$, where $\boldsymbol{\mu}$ is the mean direction, and κ stands for the concentration parameter. Therefore, the mutation operator can be perceived as a two stage process: first the direction of mutation is chosen according to $\mathcal{M}(\boldsymbol{\mu}, \kappa)$, and then the phenotype of an individual is changed in this direction by adding an one dimensional generate variable (in fact, heavy-tailed symmetric α-stable distribution $S_\alpha S(\sigma)$ is utilized [14] i.e.:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + r\boldsymbol{d}_k, \tag{5}$$

where $r \stackrel{d}{=} \chi_{\alpha,\sigma} = |S_\alpha S(\sigma)|$, and $\boldsymbol{d}_k \stackrel{d}{=} \mathcal{M}(\boldsymbol{\mu}, \kappa)$.

The direction distribution simulation process is summarized in the Tab. 1.

The concept of directional distributions presented in the previous section allows to remove the above-mentioned problem. Naturally, the effectiveness of EAs with mutation based on a class $\mathcal{M}(\boldsymbol{\mu}, \kappa)$ will depend at least on two factors: the correctness of establishing the mean direction of mutation $\boldsymbol{\mu}$, and the value of the concentration parameter κ, which controls the dispersion around the mean direction. In fact, concentration parameter κ allows to obtain on one side an isotropic distribution on the sphere ($\kappa = 1$), and on the other side, a degenerate distribution at the mean direction ($\kappa = 0$). The idea of forcing mutation direction boils down to utilizing a traditional way of creating a new individual (5). Since directional distributions, are parameterized by a pair $\{\boldsymbol{\mu}, \kappa\}$, then one must determine the strategy of adjusting their values. In the literature, several techniques of adjusting $\boldsymbol{\mu}$ can be found [12]. In this work the approach proposed by Obuchowicz [8] for the class of evolutionary algorithms known as *Evolutionary Search with Soft Selection and Forced Direction of Mutation* (ESSS-FDM):

$$\boldsymbol{\mu}^t = \frac{\langle \boldsymbol{x}^t \rangle - \langle \boldsymbol{x}^{t-1} \rangle}{\|\langle \boldsymbol{x}^t \rangle - \langle \boldsymbol{x}^{t-1} \rangle\|}, \quad \text{where} \quad \langle \boldsymbol{x}^t \rangle = \frac{1}{\eta}\sum_{k=1}^{\eta} \boldsymbol{x}_k^t$$

Table 1. Algorihtm to simulating directional distribution $\mathcal{M}(\boldsymbol{\mu}, \kappa)$

Input data

$\boldsymbol{\mu} \in \mathbb{R}^n$ – mean direction

$\kappa \in (0, 1]$ – concentration parameter

Output data

Y – pseudo-random vector of $\mathcal{M}(\boldsymbol{\mu}, \kappa)$ distribution

Algorithm

$t = 2\beta(\frac{n-1}{2}, \frac{\kappa(n-1)}{2}) - 1$, where $\beta(a, b)$ gives random number from Beta distribution

$X \leftarrow \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{n-1})$

$Z \leftarrow X/\|X\|_2$

$Y \leftarrow [\sqrt{1-t^2}Z^T, t]^T$

$Y \leftarrow [\boldsymbol{I}_n - \boldsymbol{v}\boldsymbol{v}^T]\, Y$ where $\boldsymbol{v} = \frac{[0,0,\ldots,1]^T - \boldsymbol{\mu}}{\|[0,0,\ldots,1]^T - \boldsymbol{\mu}\|_2}$

---

is chosen.

3 Experiments

3.1 Experiments description

Evolutionary algorithms used in simulation experiments in this work are based on a probably the simplest selection-mutation model of the Darwinian's evolution. The evolution is a motion of individuals in the phenotype space, called also the adaptation landscape. This motion is caused by the selection and mutation process. Selection leads to the concentration of individuals around the best ones, but mutation introduces the diversity of phenes and disperses the population in the landscape. As a mutation operator the directional mutation (5) based on the α-stable distribution is chosen and the tournament selection is chosen as a reproduction process.

The above-described algorithm is controlled by six parameters: the population size η and the maximal number of iterations $t_{\max}$, the stability index α, the slope σ and the concentration κ parameters for mutations, and the tournament size λ for selection. Parameters used in experiments are: $\alpha = 0.5, 1, 1.5, 2$, $\sigma = 0.1$, $\kappa = 0.01, 0.1, 0.5, 1$, $\eta = 10, 20, 40, 80$, $\lambda = 2, 4, 8$.

Eight benchmark functions are chosen for simulation experiments (dimensions $n = 2, 4, 8, 16$):

- five unimodal functions: Sphere model, Generalized Rosenbrock's, Schwefel's 2.21, Schwefel's 1.2, Schwefel's 2.22 functions;
- three multimodal functions: Ackley's, Generalized Rastringin's and Generalized Griewank's functions.

There have been 100 algorithm processing for each set of parameters (768 sets) for each objective function (it gives over 630000 searching processes).

3.2 Results

Due to the limited length of this paper and large size of calculated results, which are still in processed and analyzed, the result presentation will be restricted to a set of observations and remarks.

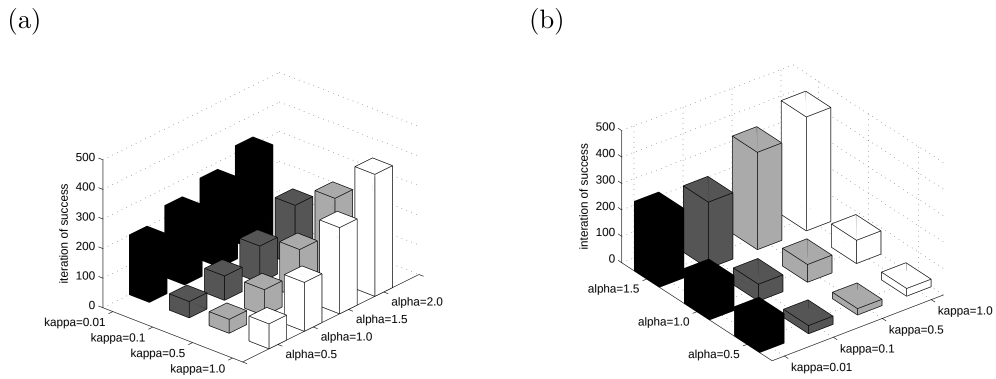

Figure 1. The mean number of iteration needed to locate the optimum with a given accuracy vs. the concentration parameter and stability index; (a) 4D sphere function, $\eta = 20$, $\lambda = 8$, success: fitness lower than 0.05; (b) 4D Rastringin's function, $\eta = 20$, $\lambda = 4$, success: fitness lower than 0.5, most algorithms runs for $\alpha = 2.0$ do not localize optimum in $t_{\max} = 15000$ iterations. Results averages over 100 algorithms runs for each set of parameters.

Dependency on the concentration parameter κ – unimodal cases: The main conclusion, which can be formulated basing on all results for unimodal functions (Figure 1(a) presents results for the sphere model), is that the worst result of the local optimization problem are obtained for extreme values of $\kappa = 0.01, 1.0$. In the case of $\kappa = 0.01$ the direction of mutation is too close to the *promising* direction $\boldsymbol{\mu}$. In fact the mutation is done almost exactly in one dimension. In the case of $\kappa = 1.0$ we randomly choose direction with the uniform distribution over the unit sphere. For all considered fitness functions, obtained results point at $\kappa = 0.1$ (from the set $\kappa = 0.01, 0.1, 0.5, 1.0$) as the best measure of the dispersion of mutation direction. This result is independent, almost without exceptions, on adjusting other input parameters.

Dependency on the concentration parameter κ – multimodal cases: Unlike the unimodal case, some correlation between the concentration parameter κ and stability index α can be observed (fig. 1(b)). In the case of low values of α, the efficiency of the evolutionary search is higher for the highest value of $\kappa = 1.0$, i.e. when the direction is chosen with uniform distribution over the unit sphere. This relation reverses with increasing α. If heavy tails of mutation distributions becomes smaller then the mutation have to be much closer to the *promising* direction $\boldsymbol{\mu}$. Unfortunately, most algorithms runs for $\alpha = 2.0$ do not localize optimum in a given limit of maximum iterations ($t_{\max} = 15000$ for each multimodal function). In the case of generalized Grewiank's and Ackley's functions algorithms with $\alpha = 0.5$ either find global optimum in several iterations or do not find it at all in a given limit of time.

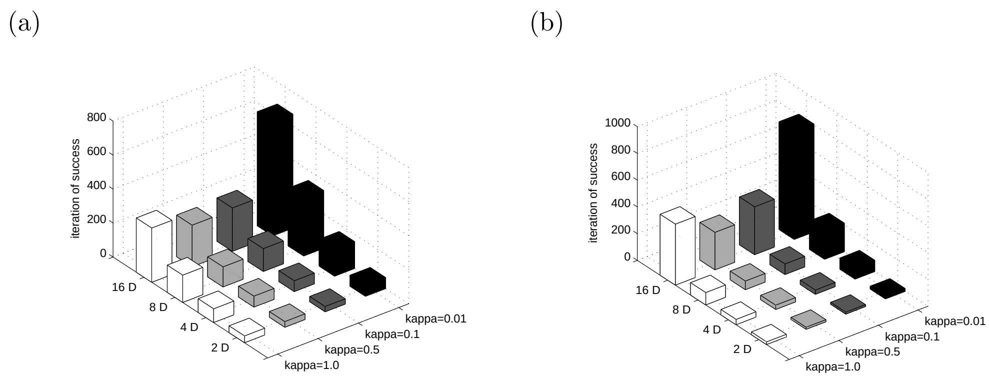

Figure 2. The mean number of iteration needed to locate the optimum with a given accuracy vs. the concentration parameter and landscape dimension; (a) sphere function, $\alpha = 1.0$, $\eta = 40$, $\lambda = 8$, success: fitness lower than 0.05; (b) Rastringin's function, $\alpha = 1.0$, $\eta = 20$, $\lambda = 4$, success: fitness lower than 0.5. Results averages over 100 algorithms runs for each set of parameters.

Dependency on the searching space dimension n: Application of the directional distribution with middle values of the concentration parameter ($\kappa = 0.1, 0.5$) to the mutation operator increases the algorithm effectiveness in problems of the local as well as the global optimum localization (fig. 2). The disproportion between this effectiveness for these values of the concentration parameter and their extreme values of ($\kappa = 0.01, 1.0$) rapidly increases with the increasing searching space dimension. This conclusion is right for unimodal as well as multimodal fitness functions considered.

Dependency on the stability index α – unimodal cases: Algorithms convergence to the local optimum points increases with the decreasing value of α. The disproportion between slopes of convergence curves decrease with increasing the concentration parameter κ and the space dimension. But, the accuracy of te local optimum point localization possesses the opposite relation. Algorithms with higher values of α converge to the optimum slower but localize it with better precision (fig. 3(a)). This fact is almost independent on the concentration parameter κ and confirms the existence of, so called, *surrounding effect* described in [9].

Dependency on the stability index α – multimodal cases: The exploration abilities are characterized by more complicated tendencies. Algorithms with extreme values of $\alpha = 0.5, 2.0$ (and the fixed slope parameter $\sigma = 0.1$) have troubles with the global optimum localization at all in the limit of iterations $t_{\max} = 15000$. The best results are obtained for $\alpha = 1.0$ (the Cauchy distribution) and $\alpha = 1.5$ (slightly worse in this case).

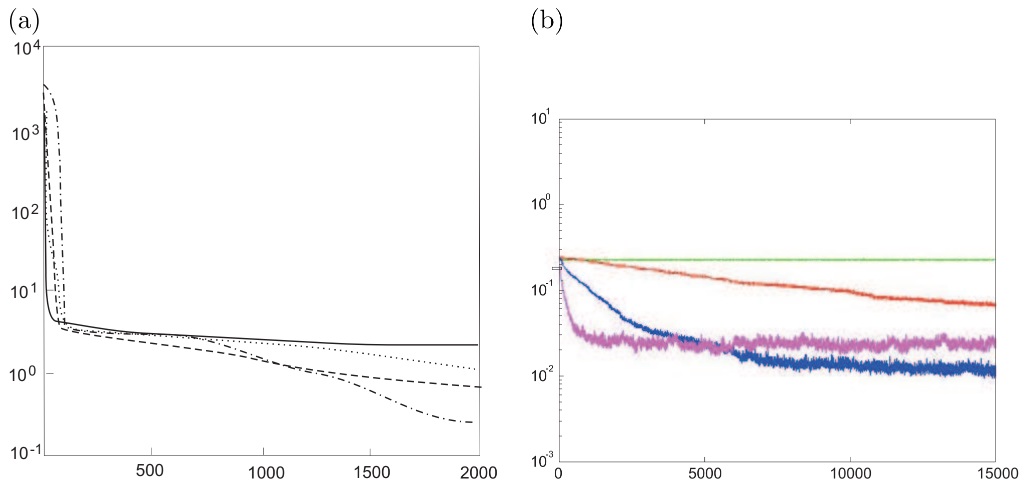

Figure 3. The best fitness in the population vs. iterations for (a) Rosenbrock's function ($\eta = 40$, $\lambda = 4$, $\kappa = 0.5$, $\alpha = 0.5$ – solid line, $\alpha = 1.0$ – dotted line, $\alpha = 1.5$ – dashed line, $\alpha = 2.0$ – dash-dott line), and (b) Grewiank's function ($\eta = 20$, $\lambda = 2$, $\kappa = 0.1$, the curve on the top for $t = 15000$ represents process with $\alpha = 2.0$, next ones – $\alpha = 1.5$ and $\alpha = 0.5$, and $\alpha = 1.0$ on the bottom). Results averages over 100 algorithms runs for each set of parameters.

The disproportion between algorithms' effectiveness for $\alpha = 1.0$ and $\alpha = 1.5$ rapidly increases with the increasing searching space dimension (fig. 2(b)). Usually, like in the unimodal case, algorithms with higher values of α converge to the optimum slower but localize it with better precision (fig. 3(b)).It is important do notice, that the influence of α selection on the global optimum localization seems to be more significant that in the case of κ adjusting. Distinct from the pair of parameters (α, σ), which adjustment processes are strongly correlated [10], processes of the selection of the best possible values of the pair (α, κ) is supposed to can be separated, because of slight correlation between these two parameters.

4 Conclusions

In this paper, the general concept for the adaptation of a probabilistic measure in a mutation operator of phenotypic EAs is considered. The proposed approach is based on the directional distributions which are parameterized by the mean vector and concentration parameter. Simulated experiments confirms the thesis that proposed mutation improves the effectiveness of evolutionary algorithms in the case of the local as well as global optimization problems. It also provides an access to construction of more sophisticated techniques which aim at improving effectiveness of evolutionary algorithms for high-dimensional problems.

Bibliography

[1] T. Bäck, D.B. Fogel and Z. Michalewicz, (Eds.). *Handbook of Evolutionary Computation*. Institute of Physics Publishing and Oxford University Press, NY, 1997

[2] I. Karcz-Dulęba. Asymptotic behaviour of a discrete dynamical system generated by a simple evolutionary process. *Int. J. Appl. Math. Comput. Sci.*, Vol. 14, No. 1, pp.79-90, 2004.

[3] K.-T. Fang, S. Kotz, and K.-W. Ng. *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London, 1990.

[4] M. Gutowski. Lévy flights as an underlying mechanism for a global optimization algorithm. Proc. 5th Conf. *Evolutionary Algorithms and Global Optimization*, Jastrzębia Gora, Poland, Warsaw University of Technology Press, 2001, pp. 79–86.

[5] Ch.Y. Lee, and X. Yao. Evolutionary Programming Using Mutations Based on the Lévy Probability Distribution. *IEEE Transactions on Evolutionary Computation,* Vol. 8, No. 1, pp. 1–13, 2004.

[6] K.V. Mardia, and P. Jupp. *Directional Statistics.* John Willey and Sons Ltd., 2nd edition, 2000.

[7] J.P. Nolan. *Stable Distributions, Models for Heavy Tailed Data*, Springer–Verlag, Berlin Heidelberg, 2003

[8] A. Obuchowicz. *Evolutionary Algorithms in Global Optimization and Dynamic System Diagnosis.* Lubuskie Scientific Society, Zielona Góra 2003.

[9] A. Obuchowicz, and P. Prętki. Phenotypic Evolution with Mutation Based on Symmetric α-Stable Distributions. *Int. J. Appl. Math. Comput. Sci.*, Vol.14, No.3, 2004, pp.289–316.

[10] A. Obuchowicz, and P. Prętki. Evolutionary algorithms with α-stable mutations. Proc. IEEE 4th International Conference on *Intelligent Systems Design and Application - ISDA 2004*. Budapest, Hungary, 2004, CD-ROM

[11] A. Obuchowicz, and P. Prętki. Isotropic Symmetric α-Stable Mutations for Evolutionary Algorithms. Proc. *IEEE Congress on Evoutionary Computation CEC'05*, Edinbourgh, UK, pp.404–410.

[12] P. Prętki, and A. Obuchowicz. Directional distributions and their application to evolutionary algorithms. *Lecture Notes in Artificial Intelligence: Artificial Intelligence and Soft Computing - ICAISC 2006*, Vol. 4029, 2006, pp.440–449.

[13] G. Rudolph. On Correlated Mutations in Evolution Strategies. Proc. 2nd Int. Conf. *Parallel Problem Solving from Nature*, Brussels 1992.

[14] G. Samorodnitsky and M.S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.