Paper

Self-Adaptive Stable Mutation Based on Discrete Spectral Measure for Evolutionary Algorithms

Andrzej Obuchowicz and Przemysław Prętki

Institute of Control and Computation Engineering, University of Zielona Góra, Zielona Góra, Poland

Abstract—In this paper, the concept of a multidimensional discrete spectral measure is introduced in the context of its application to the real-valued evolutionary algorithms. The notion of a discrete spectral measure makes it possible to uniquely define a class of multivariate heavy-tailed distributions, that have recently received substantial attention of the evolutionary optimization community. In particular, an adaptation procedure known from the distribution estimation algorithms (EDAs) is considered and the resulting estimated distribution is compared with the optimally selected referential distribution.

Keywords— discrete spectral measure, evolutionary algorithms, heavy-tailed distributions, mutation parameters adaptation.

1. Introduction

Evolutionary Algorithms (EAs) have been successfully applied to global optimization problems in many areas of engineering. Their advantage over many other optimization techniques consists in the fact that EAs are based only on function evaluations and comparisons [1]. Thus, EAs are able to deal successfully with problems that cannot be easily solved by standard optimization procedures. Unfortunately, the EAs also suffer from many serious drawbacks. The most severe one is related to the appropriate choice of their control parameters, which to a large extent determine their performance. Usually, control parameters such as a strength of a mutation, a population size, and a selective pressure are chosen during trial-and-error process or on the base of the expert knowledge, which, unfortunately, is usually inaccessible or the cost of its collection exceeds decidedly the computational cost of the optimization process itself. One way out of these difficulties is to apply algorithms, which make use of some heuristics and dedicated techniques that aim at adjusting some of the control parameters automatically during the optimization process.

In spite of the fact that the problem of the parameter adaptation has been attacked from various angles by many authors and a number of relevant results have already been reported in the literature, there is still a lack of an unified theory that addresses the problem being undertaken.

In the case of the EAs, most attention has been directed toward a normal distribution-based mutation. Thus, several relevant approaches to the adaptation of its parameters have been already reported in the literature [2], [3]. On the other hand, it is noticeable that a normal distribution does not guarantee the highest performance of EAs, so that other distributions have recently aroused evolutionary algorithms community interest. In particular, a lot of attention has been drawn to the heavy-tailed, α-stable distribution [4]–[10]. It turns out that evolutionary algorithms, which make use of the distribution of this class, gain abilities that allow them to find a balanced compromise between exploitation and exploration of the search space [11].

In general, the application of the multidimensional stable distributions to the global optimization algorithms has been limited to the simplest cases: the mutation of the base point obtained by adding a random vector composed of stable, independent, random variables [6], [7], [9], [10], or an isotropic random vector [8], [12]. This limitation causes that many properties of the stable distributions, which can turn out valuable in the context of the optimization processes, are not exploited. In order to obtain the possibility of modeling complicated dependencies between decision variables, the Discrete Spectral Measure (DSM) is proposed to generate a wide class of random vectors [13]. Unfortunately, the mutation distribution generation for a given parent solution is of a high time complexity. In [14], an estimation distribution algorithm (EDA) [15] dedicated to the evolutionary strategy $(1,\lambda)ES$ is proposed. The aim of this work is a comparative analysis of the distributions obtained by the EDAs with optimally selected referential distributions described in [13].

The paper is organized as follows. Multivariate α-stable distributions are defined in Section 2. Section 3 contains a definition and main properties of the stable random vectors based on the discrete spectral measure. In Section 4, an adaptive scheme that aims at adjusting discrete spectral measure is briefly introduced. A set of simulation experiments and their solutions are described and concluded in Section 5. Section 6 concludes the paper.

2. Univariate Stable Distribution

Definition 1: A random variable X is *stable* or *stable in the broad sense* if for X_1 and X_2 independent copies of X and any positive constants a and b,

$$aX_1 + bX_2 \stackrel{d}{=} cX + d \tag{1}$$

for some positive c and some $d \in \mathbb{R}$ and $\stackrel{d}{=}$ means that the left and right random vectors have the same distribution.

A random variable is strictly stable or stable in the narrow sense if Eq. (1) holds with $d = 0$ for all choices of a and b.

Due to the lack of the closed form formulas for densities, the stable distribution can be most conveniently described by its characteristic function (ch.f.) $\varphi(k)$ – the inverse Fourier transform of the probability density function (pdf). The ch.f. of the stable distribution is parameterized by a quadruple $(\alpha,\beta,\sigma,\mu)$ [16], where α $(0<\alpha\le 2)$ is a stability index (tail index, tail exponent or characteristic exponent), β $(-1\le\beta\le 1)$ is a skewness parameter, σ $(\sigma>0)$ is a scale parameter and μ is a location parameter. There are a variety of formulas of the ch.f. of the stable distribution in the relevant literature. This fact is caused by a combination of the historical evolution and numerous problems that have been analyzed using specialized forms of them. The most popular formula of the ch.f. of $X\sim S_\alpha(\beta,\sigma,\mu)$, i.e., a *α-stable random variable* (called also *Lévy-stable* or *just stable*) with parameters α, β, σ and μ, is given by [16]:

$$\varphi(k)=\exp\left(-\sigma^\alpha|k|^\alpha\left\{1-i\beta\,\mathrm{sign}(k)\tan\left(-\frac{\pi\alpha}{2}\right)\right\}+i\mu k\right), \quad (2)$$

when $\alpha\neq 1$, and

$$\varphi(k)=\exp\left(-\sigma|k|\left\{1+i\beta\,\mathrm{sign}(k)\frac{2}{\pi}\log|k|\right\}+i\mu k\right), \quad (3)$$

when $\alpha = 1$.

In a general case, the complexity of the problem of simulating sequences of α-stable random variables results from the fact that there is no an analytical form for the inverse of the cumulative distribution function (cdf) away from the Gaussian distribution $S_2(0,\sigma,\mu)$, Cauchy distribution $S_1(0,\sigma,\mu)$, and Lévy distribution $S_{1/2}(1,\sigma,\mu)$. The first breakthrough was made by Kanter [17], who gave a direct method for simulating $S_\alpha(1,1,0)$ random variables, for $\alpha<1$. In general cases the following result of Chambers, Mallows and Stuck [18] gives a method for simulating any α-stable random variable [7], [19].

Theorem 1: Let V and W be independent with $V\sim U(-\frac{\pi}{2},\frac{\pi}{2})$, W exponentially distributed with mean 1, $0<\alpha\le 2$.

1. The symmetric random variable

$$Z=\begin{cases}\frac{\sin(\alpha V)}{(\cos(V))^{1/\alpha}}\left[\frac{\cos((\alpha-1)V)}{W}\right]^{(1-\alpha)/\alpha} & \alpha\neq 1,\\ \tan(V) & \alpha=1\end{cases}$$

has an $S_\alpha(0,1,0)=S\alpha S$ distribution.

2. In the nonsymmetric case, for any $-1\le\beta\le 1$, define $B_{\alpha,\beta}=\arctan(\beta\tan(\pi\alpha/2))/\alpha$ when $\alpha\neq 1$. Then

$$Z=\begin{cases}\frac{\sin(\alpha(B_{\alpha,\beta}+V))}{(\cos(\alpha B_{\alpha,\beta})\cos(V))^{1/\alpha}}\left[\frac{\cos(\alpha B_{\alpha,\beta}+(\alpha-1)V)}{W}\right]^{(1-\alpha)/\alpha} & \alpha\neq 1,\\ \frac{2}{\pi}\left[\left(\frac{\pi}{2}+\beta V\right)\tan(V)-\beta\ln\left(\frac{\frac{\pi}{2}W\cos(V)}{\frac{\pi}{2}+\beta V}\right)\right] & \alpha=1\end{cases}$$

has an $S_\alpha(\beta,1,0)$ distribution.

It is easy to get V and W from independent uniform random variables $U_1,U_2\sim U(0,1)$: set $V=\pi(U_1-\frac{1}{2})$ and $W=-\ln(U_2)$. Given the formulas for the simulation of standard α-stable random variables (Theorem 1), an α-stable random variable $X\sim S_\alpha(\beta,\sigma,\mu)$ for all admissible values of the parameters α, β, σ and μ has the form

$$X=\begin{cases}\sigma Z+\mu & \alpha\neq 1,\\ \sigma Z+\frac{2}{\pi}\beta\sigma\ln(\sigma)+\mu & \alpha=1,\end{cases} \quad (4)$$

where $Z\sim S_\alpha(\beta,1,0)$.

Many interesting properties of α-stable distributions the reader can find in [7], [11], [20].

3. Multivariate Stable Distribution

There are many alternative and equivalent ways to define stable multivariate distributions [20]. One of them is based on the form of their ch.f.

Definition 2: The ch.f. $\varphi(\mathbf{k})=E[\exp(-i\mathbf{k}^T\mathbf{X})]$ of the random stable vector $\mathbf{X}$ has the following form:

$$\varphi(\mathbf{k})=\exp\left(-\int_{S^{(d)}}|\mathbf{k}^T\mathbf{s}|^\alpha\left(1-i\,\mathrm{sign}(\mathbf{k}^T\mathbf{s})\tan\left(\frac{\pi\alpha}{2}\right)\right)\Gamma(d\mathbf{s}) + i\mathbf{k}^T\mu\right) \quad (5)$$

for $\alpha\neq 1$, and

$$\varphi(\mathbf{k})=\exp\left(-\int_{S^{(d)}}|\mathbf{k}^T\mathbf{s}|\left(1-i\frac{2}{\pi}\mathrm{sign}(\mathbf{k}^T\mathbf{s})\ln|\mathbf{k}^T\mathbf{s}|\right)\Gamma(d\mathbf{s}) + i\mathbf{k}^T\mu\right) \quad (6)$$

for $\alpha=1$, where $\Gamma(\cdot)$ is the so-called spectral measure, $\boldsymbol{\mu}$ stands for shift vector, and

$$\mathrm{sign}(x)\begin{cases}1 & \text{if} \quad x>0,\\ 0 & \text{if} \quad x=0,\\ -1 & \text{if} \quad x<0.\end{cases} \quad (7)$$

It turns out that a pair $\{\Gamma,\boldsymbol{\mu}\}$ uniquely determine a stable distribution [20]. It is worth to notice that any linear combination of the components of the stable vector described by Definition 2 is univariate α-stable variable $S_\alpha(\beta,\sigma,\mu)$. It must to stressed that the definition of stable vectors is not straightforward and the presence of spectral measure Γ causes that the class is not an ordinary parametric family. In consequence, a direct definition in practical applications is used rather occasionally. Indeed, in the subsequent part of the paper, our attention is restricted only to the class of stable distributions with discrete spectral measure which possess decidedly simpler form.

4. Stable Distributions with Discrete Spectral Measure

A DSM Γ can be defined by means of Delta Dirac distribution in the following way:

$$\Gamma(\cdot;\boldsymbol{\xi},\boldsymbol{\gamma})=\sum_{i=1}^{n_s}\gamma_i\delta_{\mathbf{s}_i}(\cdot), \quad (8)$$

where $\boldsymbol{\xi} = \{\mathbf{s_i}\}_{i=1}^{n_s}, \mathbf{s_i} \in \partial S^{(d)}$ is a set of support points concentrated on a surface of a d-dimensional unit sphere, and $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^{n_s}, \gamma_i \in \mathbb{R}_+$ stands for the set of their weights. In this way, for every set $A \subset \partial S^{(d)}$ its measure is given by:

$$\Gamma(A) = \sum_{i=1}^{n_s} \gamma_i I_A(\mathbf{s}_i), \tag{9}$$

where $I_A(\cdot)$ is an indictor function of the set A. Characteristic function (5) and (6) in the case of spectral measure Eq. (8) has the form [19]:

$$\varphi(\mathbf{k}) = \exp\left(-\sum_{i=1}^{n_s} \gamma_i |\mathbf{k}^T \mathbf{s_i}|^\alpha \left(1 - i\,\mathrm{sign}(\mathbf{k}^T \mathbf{s_i}) \tan\left(\frac{\pi\alpha}{2}\right)\right) + i\mathbf{k}^T \boldsymbol{\mu}\right) \tag{10}$$

for $\alpha \neq 1$, and

$$\varphi(\mathbf{k}) = \exp\left(-\sum_{i=1}^{n_s} \gamma_i |\mathbf{k}^T \mathbf{s_i}| \left(1 - i\frac{2}{\pi}\mathrm{sign}(\mathbf{k}^T \mathbf{s_i}) \ln|\mathbf{k}^T \mathbf{s_i}|\right) + i\mathbf{k}^T \mu\right) \tag{11}$$

for $\alpha = 1$.

The definition of the DSM allows to use multivariate stable distributions in the simpler way. It is worth to notice that application of the DSM does not limit any properties of multivariate stable vectors. The following theorem can be proved [21].

Theorem 2: Let $p(\mathbf{x})$ be a density function of the stable distribution described by the characteristic function (5) and (6), and $p^*(\mathbf{x})$ is a density function of the random vector described by the characteristic function (10) and (11), then

$$\begin{aligned} &\forall \varepsilon > 0 \quad \exists n_s \in \mathbb{N} \quad \exists \boldsymbol{\xi}, \gamma \quad \forall \mathbf{x} \in \mathbb{R}^d : \\ &\sup_{\mathbf{x}\in\mathbb{R}^d} |p(\mathbf{x}) - p^*(\mathbf{x})| < \varepsilon. \end{aligned} \tag{12}$$

In other words, each stable distribution can be approximated by some distribution based on the DSM with any accuracy. Especially, the existence of a procedure of pseudo-random vectors generation is very important. It turns out, that a simulation procedure of stable random vectors $\mathbf{X}$ defined by the characteristic functions (10) and (11) is straightforward, and can be implemented by the following stochastic decomposition:

$$\mathbf{X} \overset{d}{=} \begin{cases} \sum_{i=1}^{n_s} \gamma_i^{1/\alpha} Z_i \mathbf{s}_i & \text{for} \quad \alpha \neq 1, \\ \sum_{i=1}^{n_s} \gamma_i^{1/\alpha} \left(Z_i - \frac{2}{\pi}\ln(\gamma_i)\right) \mathbf{s}_i & \text{for} \quad \alpha = 1, \end{cases} \tag{13}$$

where Z_i are i.i.d. stable random variables $S_\alpha(1,1,0)$ for which an effective generator can be found in [20].

Random vectors ($\mathbf{X} = [X_1, X_2, \ldots, X_n], X_i \sim S_\alpha S(\sigma)$) composed of independent symmetric elements possess a special status in the application to mutation operators of EAs [5], [6], [7], [14], [9], [10]. It occurs that random vectors can be enriched by adding μ and β parameters, i.e., $\mathbf{X} = [X_1, X_2, \ldots, X_n], X_i \sim S_\alpha(\sigma, \beta, \mu)$ if the DSMs are applied. It means that each component acquires additional degrees of freedom. This fact is very important in the context of application of the above random vector to modeling complicated dependencies between decision variables. It is easy to show that an exploration of such dependencies and their inclusion to a mutation operator accelerates the optimization process. In order to illustrate the possibilities of the DSM representation of the random vectors, it can be mention that the vector with independent components $X_i \sim S_\alpha S(\sigma)$ have the DSM focused in the points of orthogonal axes and surface of the unit sphere intersection with a different weights. The versatility of the DSM representation of the distribution is included in [20].

Theorem 3: The spectral measure of the stable vector $\mathbf{X}$ is described by a finite number of the support vectors $\mathbf{s}_i$ if, and only if the vector $\mathbf{X}$ can be represented by a linear combination of the independent stable random variables, i.e.:

$$\mathbf{X} = \mathbf{AZ}, \tag{14}$$

where $\mathbf{A} \in \mathbb{R}^{d\times N}$, $\mathbf{Z} = [Z_1, \ldots, Z_N]^T, Z_i \sim S_\alpha(\sigma, \beta, \mu)$.

Based on the Theorem 3, it can be shown that the DSM can be also applied to represent vectors which are described by parameters σ, β, μ and by stochastic dependencies between these vectors.

One of the important properties of the stable distribution based on the DSM and a finite set of support vectors is that almost all probability mass remote from the base point is focused around directions described by the support vectors. So, macromutations take place only in direction parallel to the DSM support vectors. This effect is illustrated in Fig. 1.

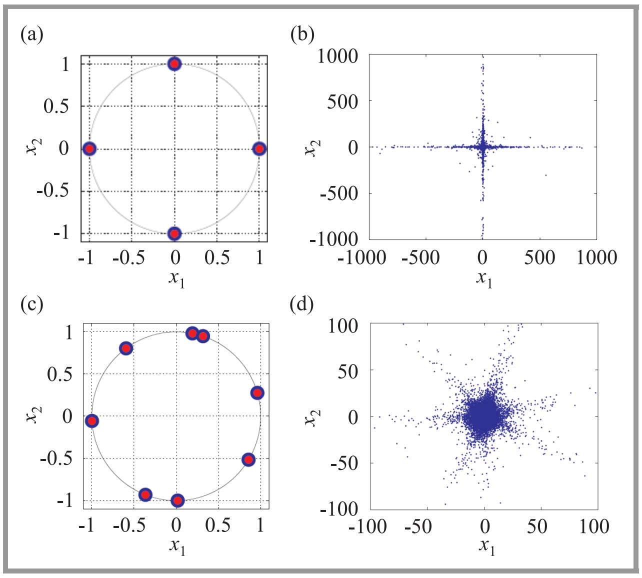

Fig. 1. Distribution of the DSM support vectors and corresponded random realizations: (a), (b) – α=0.75, (c), (d) – α=1.5.

Summarizing, two main benefits obtained by the application of the DSM representation of a multivariate stable distribution can be distinguished: macromutations which allows simple cross saddles of the searching environments, and possibility of modeling complex stochastic dependencies. The above mentioned benefits are illustrated in [13].

5. Learning Probability Model with Discrete Spectral Measure

One of the most important factors influencing an effectiveness of a global optimization procedure is the possibility of a configuration parameters adaptation. Especially, this mechanism can be applied to mutation operators in evolutionary algorithms. There are many instances in literature, in which such procedures are proposed for mutation parameters [8], [22]–[24], however, all of them consider the Gaussian mutation, only. In the case of the whole class of multidimensional stable distributions, Rudolph shows that adaptation procedures should be different for different stable indices α [8]. In this paper an original parameter adaptation procedure of the mutations based on the multidimensional stable distributions described by the DSM is proposed.

Let us focus our attention on the evolution strategy $(1,\lambda)ES$. In this strategy, the population of descendants is generated by λ mutations of a given base point. All descendant points are evaluated, i.e., the optimized fitness function is calculated in these points. The descendant with the best fitness is chosen as a new base point. The above described operations are iteratively repeated until a given stop criterion is met.

The necessity of adjusting probability model of the mutation utilized to explore a search space is undisputed. It can be noticed that a fixed probability distribution usually causes many serious problems. On one hand, tails of the distribution might be too narrow to allow an algorithm to escape from the local solution in a reasonable number of generations. On the other hand, the probability mass focused around the center of the mutation might be insufficient to make a significant progress in improving an estimate of the global solution. An ideal adaptation procedure should be able to detect each of these situations, and adjust probabilistic model in such a way to prevent the algorithm from a stagnation.

The method of the optimal probabilistic model choice is introduced and illustrated in [13]. The goal is the correction of the current solution by a perturbation using a stable random vector $\mathbf{X}^{\boldsymbol{\gamma}}_{\boldsymbol{\xi}}$. The aim of this calculation is the selection of the optimal stable model from the class of multivariate distributions described by the DSM. If we assume the fixed set of n_s support points $\boldsymbol{\xi}$, the criterion of the best model selection is chosen in the form:

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{n_s}_+} C(\boldsymbol{\gamma}), \tag{15}$$

where

$$C(\boldsymbol{\gamma}) = E\left[\min\left\{\frac{\phi(\mathbf{x}_k + \mathbf{X}^{\boldsymbol{\gamma}}_{\boldsymbol{\xi}}(\alpha))}{\phi(\mathbf{x}_k)}, 1\right\}\right]. \tag{16}$$

It occurs that the function (16) does not possess an analytical form, thus, the problem Eq. (15) cannot be solved using standard optimization techniques. One of the possible solutions is the application of the Monte Carlo method [25]. The law of large numbers [26] allows to approximate the expectation value (16) using the following estimator:

$$\hat{C}(\boldsymbol{\gamma}) = \frac{1}{N}\sum_{i=1}^{N}\min\left\{\frac{\phi(\mathbf{x}_k + \mathbf{X}^{\boldsymbol{\gamma}}_{i,\xi}(\alpha))}{\phi(\mathbf{x}_k)}, 1\right\}, \tag{17}$$

where $\{\mathbf{X}^{\boldsymbol{\gamma}}_{i,\boldsymbol{\xi}}(\alpha)\}_{i=1}^N$ stands for a sequence of independent realizations of the random vector with the α-stable distribution. Using the estimator (17), the problem Eq. (15) can be rewritten into the form:

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}\in\mathbb{R}^{n_s}} \hat{C}(\boldsymbol{\gamma}). \tag{18}$$

Because the objective function possesses a stochastic properties and in order to achieve the compromise between the computation complexity and the estimator quality, the SPSA algorithm [27] can be chosen for solving Eq. (18).

The quality of the estimator (17) strongly depends on the number N of independent realizations of the random vectors $\{\mathbf{X}^{\boldsymbol{\gamma}}_{i,\boldsymbol{\xi}}(\alpha)\}_{i=1}^N$. Experiments shows that, in order to obtain a representative estimator $\hat{C}$ (17) in the 2D searching space, N should be up to dozens thousands. Because of such a large complexity of the optimal probabilistic model choice, an adaptive scheme known from the class of the estimation of distribution algorithms (EDAs) [15] is proposed to adjust a parametric probability model Eq. (8) [14]. The idea of the proposed algorithm is based on the assumption that the selection pressure of the evolutionary process designates the most valuable set of the independent realizations of the random vectors. This idea utilized to evolutionary strategy $(1,\lambda)$ES with the mutation operator that is based on the DSM boils down to the algorithm $(1,\lambda/\mu)ES_\alpha$ composed of the following simple steps:

Step 0: Set $k = 1$ and choose an initial guess of a global solution $\mathbf{x}_0$ and initial weights vector $\boldsymbol{\gamma}_0$, i.e.:

$$\mathbf{x}_k = \mathbf{x}_0, \quad \Gamma_k = (\boldsymbol{\xi}, \boldsymbol{\gamma}_0), \tag{19}$$

where $\boldsymbol{\xi} = \{\mathbf{s}_i\}_{i=1}^{n_s}$, $\mathbf{s}_i \in S^d$ is a fixed grid of a DSM, and $\boldsymbol{\gamma}_k = \{\gamma_i\}_{i=1}^{n_s}$ stands for a vector of weights associated with support points $\mathbf{s}_i$.

Step 1: Randomly pick a set of candidate solutions using probabilistic model Γ_k.

$$\mathbf{P}_{k,\lambda} = \{\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \ldots, \mathbf{x}_{k,\lambda}\}, \quad \mathbf{x}_{k,i} = \mathbf{x}_k + \mathbf{X}_i, \tag{20}$$

where $\mathbf{X}_i \sim \Gamma_k$ are i.i.d. random vector generated according Eq. (13).

Step 2: Select μ the best solutions from the current population $\mathbf{P}_{k,\lambda}$:

$$\mathbf{P}_{k,\mu} = \{\mathbf{x}_{k,1:\lambda}, \mathbf{x}_{k,2:\lambda}, \ldots, \mathbf{x}_{k,\mu:\lambda}\} \tag{21}$$

Step 3: Build the probability model of the selected solutions $\mathbf{P}_{k,\mu}$.

$$\gamma_k = Est(\mathbf{P}_{k,\mu}) \tag{22}$$

where $Est(\cdot)$ is an estimation procedure of the weights described in the subsequent part of this section.

Step 4: Set $\mathbf{x}_k = \mathbf{x}_{k,1:\lambda}$.

Step 5: If the stopping condition are not met, go to Step 1.

The Step 3 deserves more detailed description. In the literature [19], [28]–[30] several approaches to the estimation of the DSM from a given data set can be found. The simplest and the less computational intensive one was presented in [19]. This method is based on the, so called, empirical ch.f.:

$$\hat{\phi}(\mathbf{k}) = \frac{1}{N}\sum_{i=1}^{N} \exp(j\mathbf{k}^T\mathbf{X}_i), \tag{23}$$

where $\mathbf{X}_i$ are observed random variables realizations, which are included in the data set $\{\mathbf{x}_i\}_{i=1}^{N}$. Assuming that the DSM Γ is defined by the finite set of support vectors $\boldsymbol{\xi} = \{\mathbf{s}_i\}_{i=1}^{n_s}$ and, corresponded to them, weighs $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^{n_s}$ t.j.:

$$\Gamma = \left\{ \begin{array}{cccc} \mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{n_s} \\ \gamma_1 & \gamma_2 & \dots & \gamma_{n_s} \end{array} \right\}, \tag{24}$$

the estimation problem can be reduced to the optimization problem

$$\Gamma^* = \arg\min_{\boldsymbol{\xi},\boldsymbol{\gamma}} \|\hat{\phi}(\mathbf{k}) - \phi(\mathbf{k};\boldsymbol{\xi},\boldsymbol{\gamma})\|. \tag{25}$$

In order to the problem simplification, let us assume that the considered DSM is based on the fixed set of support vectors, which are uniformly distributed on the unique sphere surface. Than the model estimation problem (25) is reduced to the form:

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}} \|\hat{\phi}(\mathbf{k}) - \phi(\mathbf{k};\mathbf{S}_{n_s},\boldsymbol{\gamma})\|. \tag{26}$$

Searching for the precise solution of the problem Eq. (26) is connected with a very large computation effort and the application of this method seems to be unpractical. We should introduce another problem simplification. The expression (26) can be estimated using the set of testing points $\mathbf{K} = \{\mathbf{k}_i\}_{i=1}^{n_k}$:

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}} \sum_{i=1}^{n_k} \left(\hat{\phi}(\mathbf{k}_i) - \phi(\mathbf{k}_i;\boldsymbol{\xi},\boldsymbol{\gamma})\right)^2 \tag{27}$$

Let $\mathbf{I} = -[\ln\hat{\phi}(\mathbf{k}_1),\dots,\ln\hat{\phi}(\mathbf{k}_{n_k})]^T$ and

$$\boldsymbol{\Psi}(\mathbf{k}_1,\dots,\mathbf{k}_{n_k};\mathbf{s}_1,\dots,\mathbf{s}_{n_s}) = \begin{pmatrix} \psi_\alpha(\mathbf{k}_1^T\mathbf{s}_1) & \dots & \psi_\alpha(\mathbf{k}_1^T\mathbf{s}_{n_s}) \\ \vdots & \vdots & \vdots \\ \psi_\alpha(\mathbf{k}_{n_k}^T\mathbf{s}_1) & \dots & \psi_\alpha(\mathbf{k}_{n_k}^T\mathbf{s}_{n_s}) \end{pmatrix} \tag{28}$$

for

$$\psi_\alpha(u) = \left\{ \begin{array}{ll} |u|^\alpha(1 - i\,\mathrm{sgn}(u)\tan\left(\frac{\pi\alpha}{2}\right)), & \text{for} \quad \alpha \neq 1 \\ |u|(1 - i\frac{2}{\pi}\mathrm{sgn}(u)\ln\left(|u|\right)), & \text{for} \quad \alpha = 1 \end{array} \right. \tag{29}$$

than the optimal weight set is the solution of the system of equations [19]

$$\mathbf{I} = \boldsymbol{\Psi}\boldsymbol{\gamma}^*. \tag{30}$$

In order to ensure well-conditional problem (30) let us assume $n_s = n_k$ and $\mathbf{s}_i = \mathbf{k}_i$. Finally, the problem is reduced to solving the following constrained quadratic programming problem [19]:

$$\boldsymbol{\gamma}^* = \arg\min_{\boldsymbol{\gamma}\geq\mathbf{0}} ||\mathbf{c} - \mathbf{A}\boldsymbol{\gamma}||_2, \tag{31}$$

where $\mathbf{c} = [\mathrm{Re}\{I_{1:n/2}\},\mathrm{Im}\{I_{n/2+1:n}\}]^T$ is a vector containing n real values of the vector I and n its image values, $\mathbf{A} = [\mathrm{Re}\{\boldsymbol{\psi}_1^T\},\dots,\mathrm{Im}\{\boldsymbol{\psi}_n^T\}]^T$, is similarly organized matrix (28), where $\boldsymbol{\psi}_i = [\psi(\mathbf{s}_1^T\mathbf{s}_i),\psi(\mathbf{s}_2^T\mathbf{s}_i),\dots,\psi(\mathbf{s}_{n_s}^T\mathbf{s}_i)]^T \in \mathbb{C}^{n_s}$ are its rows.
The problem Eq. (31) can be solved analytically or using one of the dedicated gradient-based optimization method.

6. Experimental Simulation

In this section we experimentally try to prove, that the self-adapted DSM can improve the optimization efficiency of the evolution strategy being considered.

6.1. Experiment 1

Three versions of the evolution strategy will be analyzed:

- **A1** – $(1,\lambda)\mathrm{ES}_\alpha$ for which mutation is based on the DSM with following support vectors:

  $$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

  The weight vector $\boldsymbol{\gamma} = [\sigma/4,\sigma/4,\sigma/4,\sigma/4]^T$ is fixed during the evolutionary process.

- **A2** – $(1,\lambda/\nu)\mathrm{ES}_\alpha$ for which mutation is based on the DSM with following support vectors:

  $$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\}$$

  The initial weight vector $\boldsymbol{\gamma} = [\sigma/4,\sigma/4,\sigma/4,\sigma/4]^T$ is adapted during the evolutionary process in order to the algorithm proposed in this paper.

- **A3** – $(1,\lambda/\nu)\mathrm{ES}_\alpha$ for which mutation is based on the DSM with following support vectors:

  $$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} \right\}$$

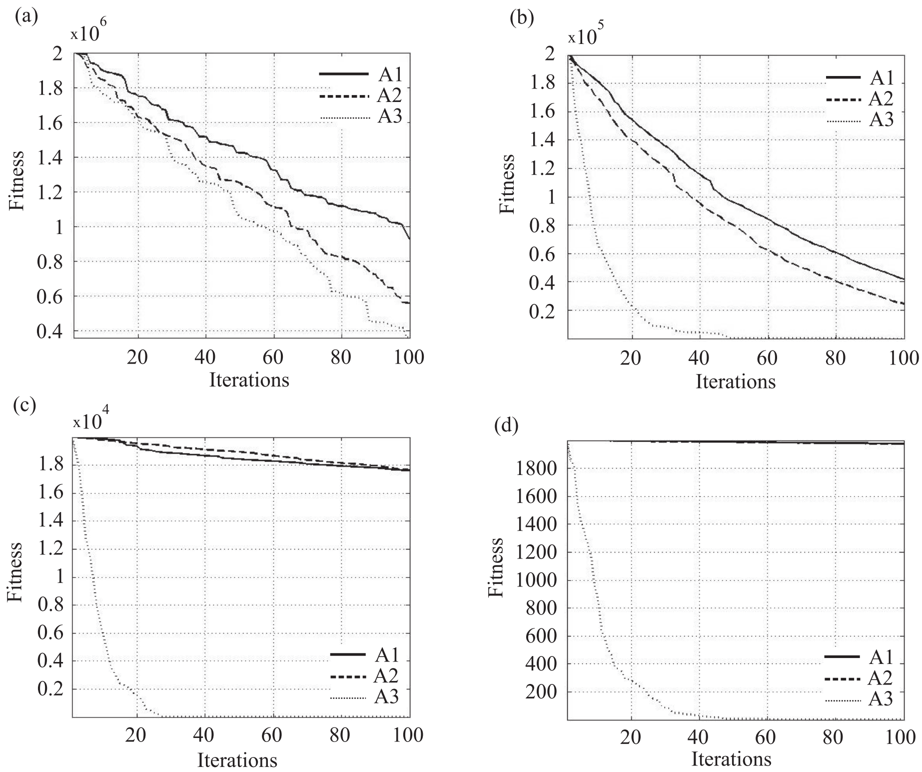

Fig. 2. Fitness of the base point versus iterations. Average results (taken over 50 algorithms independent runs) for quadratic function Eq. (32) obtained for algorithms with different DMSes: four support points without weights adaptation **A1**, four support points with weights adaptation **A2**, and eight support points with weights adaptation **A3**. Objective function parameters (e_1, e_2): (a) = $(1,1)$, (b) = $(10,0.1)$, (c) = $(100,0.01)$, (d) = $(1000,0.001)$.

The initial weight vector $\boldsymbol{\gamma} = [\sigma/8, \ldots, \sigma/8]^T$ is adapted during the evolutionary process in order to the algorithm proposed in this paper.

Four two-dimensional unimodal objective functions are selected to experiments:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbb{U}\mathbb{D}\mathbb{U}^T \mathbf{x}, \tag{32}$$

where

$$\mathbb{U} = \begin{pmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

and

$$\mathbb{D} = \begin{pmatrix} e_1 & 0 \\ 0 & e_2 \end{pmatrix},$$

for different conditional factors: $(e_1, e_2) = (1,1)$, $(10, 0.1)$, $(100, 0.01)$, $(1000, 0.001)$. The initial conditions are the same for each strategy: $\lambda = 20$, $\nu = 10$, $\mathbf{x}_0 = [1000, 1000]^T$, $\sigma = 1$. Moreover, we assume that $\sum_{i=1}^{n_s} \gamma_i = \sigma$.
The results for the algorithms with the stability index $\alpha = 0.5$ are presented in Fig. 2.
By analyzing average results presented in Fig. 2 for evolutionary strategies **A1** and **A2** the advantage of the applied adaptation mechanism cannot be declared explicitly. The choice of the stable index $\alpha = 0.5$, for which macromutations in directions parallel to the axis of the reference frame take place an important role in the optimum finding processs. Only after a condensation of the support vectors strategy **A3** there is possibility to fit the exploration distribution to any distribution of the searching space. In cases presented on Fig. 2(b)–2(d) it is easy to observe that the weight $\gamma_6 \approx \sigma$ (it is connected to the vector $\mathbf{s}_6 = [-\sqrt{2}/2, \sqrt{2}/2]$), and the macromutations in this direction are preferred.

6.2. Experiment 2

The goal of the experiments is to recognize how "far-off" is the probabilistic model Eq. (31) applied in $(1, \lambda/\mu)ES_\alpha$ from the optimal one Eq. (18). As a measure of this "distance" in the parent point $\mathbf{x}_k$ obtained in the k-th iteration of evolutionary process the following expression is chosen

$$J(\mathbf{x}_k) = \frac{E_{\Gamma_k^\mu}}{E_{\Gamma_k}}, \tag{33}$$

where

$$E_{\Gamma_k^\mu} = \int_D \min\left(f(\mathbf{x}), f(\mathbf{x}_k)\right) d\Gamma_k^\mu, \tag{34}$$

$$E_{\Gamma_k} = \int_D \min\left(f(\mathbf{x}), f(\mathbf{x}_k)\right) d\Gamma_k, \tag{35}$$

and $f(\mathbf{x})$ $(\mathbf{x} \in D)$ is an optimized fitness function, Γ_k^μ and Γ_k are cumulative distribution functions (cdf) of probability models Eqs. (31) and (18), respectively. The expression

$\min(f(\mathbf{x}), f(\mathbf{x}_k))$ in Eq. (34) is introduced instead of the pure fitness function $f(\mathbf{x})$ in order to avoid the possibility to obtain infinite expectation value of $f(\mathbf{x})$.

As it can be seen, the measure $J(\mathbf{x}_k)$ Eq. (33) is equal to the unity in the case of the perfect match of both distributions, and increases when the disproportion between both distributions increases. The value of $J(\mathbf{x}_k)$ also constitutes the measure of deterioration (in the probabilistic sense) of the next population quality when we use the probabilistic model Γ_k^{μ} Eq. (31) instead of the optimal one Γ_k Eq. (18).

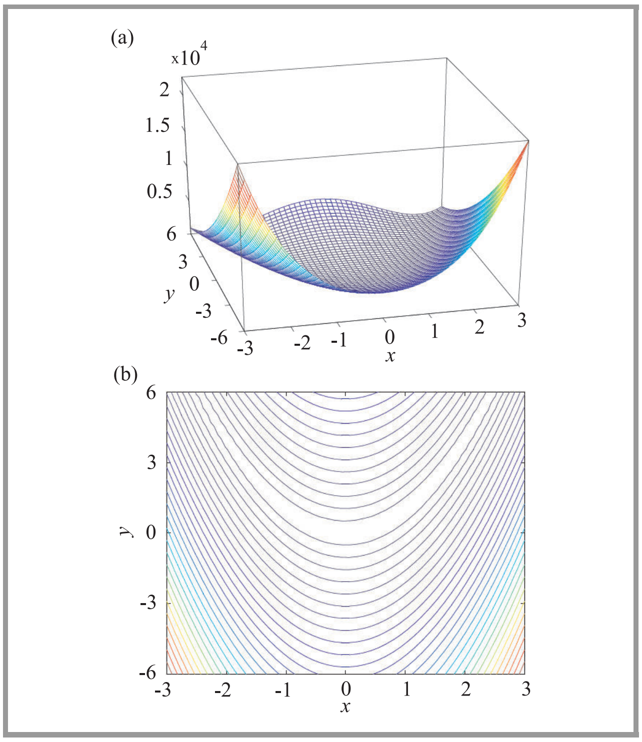

Fig. 3. Rosenbrock's function – (a) and its contour chart – (b).

Let the 2-dimensional Rosenbrock's function (Fig. 3)

$$f(\mathbf{x}) = (1-x_1)^2 + 100(x_2 - x_1^2)^2 \qquad (36)$$

be chosen as a objective function for considered computation example. Moreover, we assume that the point $\mathbf{x}_0 = [10, 10]^T$ $(f(\mathbf{x}_k) = 810081)$ is the initial approximation of the optimum. Let us reduce the set of rival probabilistic models to a set of four stable distributions $\Omega = \{\mathbf{X}_{\boldsymbol{\xi}}^{\boldsymbol{\gamma}}(\alpha) | \alpha = 0.5, 1.0, 1.5, 2.0\}$. Each random vector $\mathbf{X}_{\boldsymbol{\xi}}^{\boldsymbol{\gamma}}(\alpha)$ is described by the DSM spread on 5 uniformly distributed support points:

$$\boldsymbol{\xi} = \left\{ \begin{bmatrix} 1.0000 \\ 0.0000 \end{bmatrix}, \begin{bmatrix} 0.3090 \\ 0.9511 \end{bmatrix}, \begin{bmatrix} -0.8090 \\ 0.5878 \end{bmatrix}, \begin{bmatrix} -0.8090 \\ -0.5878 \end{bmatrix}, \begin{bmatrix} 0.3090 \\ -0.9511 \end{bmatrix} \right\}.$$

In the case of the $(1, \lambda/\mu)ES_\alpha$ algorithm the parameter $\lambda = 100$ and three values of μ are selected $(10, 40, 70)$. There are 128 algorithm runs for each pair (α, μ). In order to estimate the optimal probability model $N = 100000$ independent realizations of the random vectors $\{\mathbf{X}_{i,\boldsymbol{\xi}}^{\boldsymbol{\gamma}}(\alpha)\}_{i=1}^{N}$ is generated.

The obtained results are presented in the form of histograms of $J(\mathbf{x}_k)$ Eq. (33) for each pair (α, μ) taken over 2560 sample points (Fig. 4). It can be observed that there are distinct peaks near $J(\mathbf{x}_k) = 1$ for all cases. It suggests that proposed probability model of mutation Eq. (31) is a good estimator of the optimal probability model. The relation between the quality of this estimator and the stability index α is very interesting. The peaks of $J(\mathbf{x}_k)$ histograms are higher for extreme values of stability index $\alpha = 0.5$ and $\alpha = 2$ than for the medium values $\alpha = 1$ and $\alpha = 1.5$. The explanation of this fact is not quit clear and needs more precise research. Taking into account obtained results of our experiment, the relation between the quality of the estimator and the parameter μ is not clear. Two reasons can influence this fact. The first one, highly probable in our opinion, is that: too low number of points is considered and a statistical error is too high. The second reason can be connected with the fact that the quality of the estimator at a given point $\mathbf{x}$ depends not only on the pair (α, μ) but also on the features of the fitness landscape in the close neighborhood of $\mathbf{x}$.

7. Conclusions

The application of the discrete spectral measure to the stable mutation for evolutionary algorithms based on the real-valued representation of the individual is considered. The evolution strategy $(1+\lambda)ES_\alpha$ is chosen as a base evolutionary algorithm. Emphasis is focused on the self adaptation of the mutation probability model to the winner individual in each population. This self adaptation can be parted into two steps: the optimal selection of the set of support points (vectors) and the optimal selection of the weights related to this points. Both tasks are connected with the calculations of a high time and space complexity cost [13], [14]. This fact limits applicability of the proposed method to low dimensional problems. In order to avoid this cost the estimation methods for both tasks should be proposed. In this paper we focus on the second task, i.e., we assume a fixed set of support points. This work contains preliminary results of simulation experiments. Results suggest that probability model of the mutation is a good estimator of the optimal one. The analysis of the relation between the estimator quality and control parameters (α, μ) needs further thorough investigations. Our further research will also be focused on the support points selection method of the lower complexity cost.

References

[1] 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. 3, pp. 79–90, 2004.

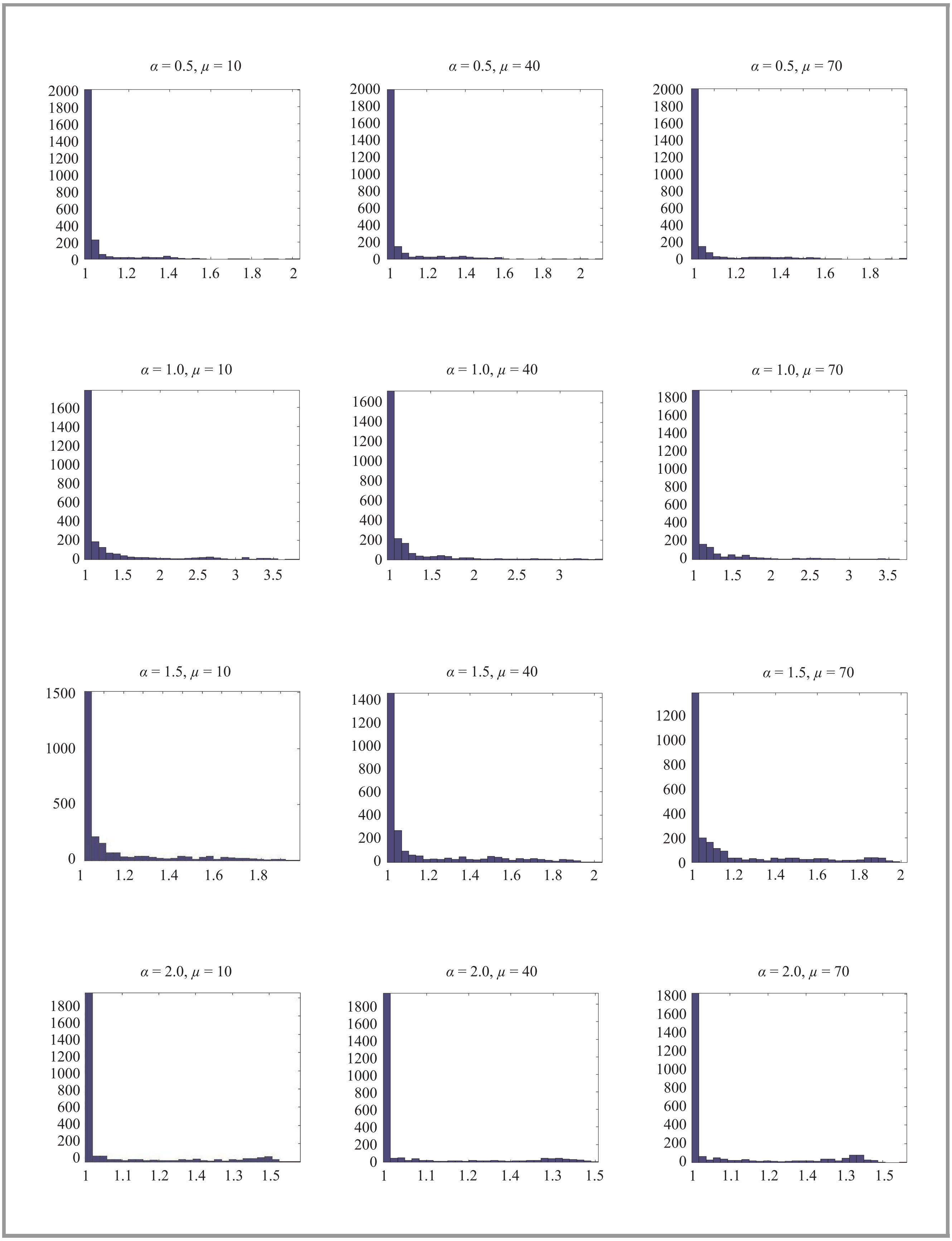

Fig. 4. Histograms of $J(\mathbf{x}_k)$ Eq. (33) for each pair (α, μ) taken over 2560 sample points.

[2] H. G. Beyer and H. P. Schwefel, "Evolutionary strategies – a comprehensive introduction", *Neural Computing*, vol. 1, no. 1, pp. 3–52, 2002.

[3] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolutionary strategies", *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, 2002.

[4] N. Hansen, F. Gemperle, A. Auger, and P. Koumoutsakos, "Why do heavy-tail distributions help?" in *Parallel Problem Solving from Nature PPSN IX*, Springer, LNCS 4193, pp. 62–71, 2006.

[5] X. Liu and W. Xu, "A new filled function applied to global optimization", *Comput. Oper. Res.* vol. 31, pp. 61–80, 2004.

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

[7] 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, pp. 289–316, 2004.

[8] G. Rudolph, "Local convergence rates of simple evolutionary algorithms wich Cauchy mutations", *IEEE Trans. Evol. Comput.* vol. 1, no. 4, pp. 249–258, 1997.

[9] X. Yao and Y. Liu, "Fast evolution strategies", in *Evolutionary Programming IV*, P. J. Angeline, R. G. Reynolds, J. R. MacDonnell, and R. Eberhart, Eds. Berlin: Springer, 1997, pp. 151–161.

[10] X. Yao, Y. Liu, and G. Liu, "Evolutionary programming made faster", *IEEE Trans. Evol. Comput.* vol. 3, no. 2, pp. 82–102, 1999.

[11] P. Prętki, "Algorytmy ewolucyjne z mutacją α-stabilną w zadaniach globalnej optymalizacji parametrycznej". Ph.D. thesis, University of Zielona Góra, Poland, 2008 (in Polish).

[12] A. Obuchowicz and P. Prętki, "Isotropic symmetric α-stable mutations for evolutionary algorithms", in *Proc. IEEE Congr. Evol. Comput. CEC'05*, Edinburgh, UK, 2005, pp. 404-410.

[13] A. Obuchowicz and P. Prętki, "Evolutionary algorithms with stable mutations based on a discrete spectral measure", in *Proc. Int. Conf. Artif. Intel. Soft Comput. ICAISC'2010*, Springer-Verlag, LNAI 6114, no. 2, pp. 181–188, 2010.

[14] P. Prętki, "Learning stable mutation in $(1,\lambda)$ES evolutionary strategy", in *Proc. 10th Conf. Evol. Algorithms Global Optimiz.*, Będlewo, Poland, 2007, pp. 233–240.

[15] P. Larranaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Optimization.* Boston: Kluwer Academic Publishers, 2001.

[16] A. Zolotariev,*One-Dimensional Stable Distributions*. Providence: American Mathematical Society Press, 1986.

[17] M. Kanter, "Stable densities under change of scale and total variation inequalities", *Ann. Probab.*, vol. 3, no. 4, pp. 697–707, 1975.

[18] J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables", *J. Amer. Statist. Assoc.*, vol. 71, pp. 340–344, 1976.

[19] J. P. Nolan, A. K. Panorska, and J. H. McCulloch, "Estimation of stable spectral measures – stable non-Gaussian models in finanse and econometrics", *Math. Comput. Model.* vol. 34, no. 9, pp. 1113–1122, 2001.

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

[21] T. Byczkowski, J. P. Nolan, and B. Rajput, "Approximation of multidimensional stable densities", *J. Mult. Anal.*, vol. 46, pp. 13–31, 1993.

[22] T. B ack, *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, 1995.

[23] I. Rechenberg, *Evolutionsstrategie: Optimierung technisher Systeme nach Prinzipien der biologishen Evolution*. Stuttgard: Frommann-Holzburg Verlag, 1973.

[24] H.-P. Schwefel, *Evolution and Optimum Seeking*. New York: Wiley, 1995.

[25] F. Kemp, "An introduction to sequential Monte Carlo methods", *J. Royal Statist. Soc.* vol. D52, pp. 694–695, 2003.

[26] R. Durrett, *Probability: Theory and Examples*. Duxbury Press, 1995.

[27] J. C. Spall, *Introduction to Stochastic Search and Optimization*. Hoboken, NJ: Wiley, 1993.

[28] L. N. Chrysostomos and M. Shao, *Signal Processing with Alpha-Stable Distribution and Applications*. Chichester, UK: Wiley, 1981.

[29] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable odeling of noise and robust time-delay in the presence of impulsive noise", *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, 1999.

[30] P. Kidmose, "Alpha-stable distributions in signal processing of audio signals", in *Proc. 41st Conf. Simulation and Modeling SIMS 2000*, Scandinavian Simulation Society Press, pp. 87–94, 2000.

Andrzej Obuchowicz received the M.Sc. and Ph.D. degrees in physics, and D.Sc. degree in automation and robotics from Technical University of Wrocław in 1987, 1992 and 2004, respectively. He is an Associate Professor in Institute of Control and Computation Engineering University of Zielona Góra, Poland. His research interests cover global optimization methods, application of pattern and image recognition techniques in medical diagnosis as well as technological system diagnosis. He is author or co-author of 2 monographs and over 100 scientific papers.
E-mail: A.Obuchowicz@issi.uz.zgora.pl
Institute of Control and Computation Engineering
University of Zielona Góra
Podgórna st 50
65-246 Zielona Góra, Poland

Przemysław Prętki received the Ph.D. degree in Computer Science from the University of Zielona Góra, Poland, in 2008. In 2003–2010 he was an Assistant Professor in the Institute of Control and Computation Engineering at the University of Zielona Góra. His research interests include methods of global optimization, machine learning, data mining, particularly mining software engineering data and automated software testing. He works now as an engineer at Advanced Digital Broadcast.
E-mail: P.Pretki@issi.uz.zgora.pl
Institute of Control and Computation Engineering
University of Zielona Góra
Podgórna st 50
65-246 Zielona Góra, Poland