Article

Spherical Distributions Used in Evolutionary Algorithms

Alexandru Agapie [1,2]

1 Department of Applied Mathematics, Faculty of Economic Cybernetics, Statistics and Informatics, Bucharest University of Economic Studies, Calea Dorobantilor 15-17, 010552 Bucharest, Romania; alexandru.agapie@csie.ase.ro
2 "Gheorghe Mihoc—Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, 050711 Bucharest, Romania

Abstract: Performance of evolutionary algorithms in real space is evaluated by local measures such as success probability and expected progress. In high-dimensional landscapes, most algorithms rely on the normal multi-variate, easy to assemble from independent, identically distributed components. This paper analyzes a different distribution, also spherical, yet with dependent components and compact support: uniform in the sphere. Under a simple setting of the parameters, two algorithms are compared on a quadratic fitness function. The success probability and the expected progress of the algorithm with uniform distribution are proved to dominate their normal mutation counterparts by order $n!!$.

Keywords: probabilistic optimization; spherical distribution; multi-variate calculus; hypergeometric functions; transition kernel

1. Introduction

Probabilistic algorithms are among the most popular optimization techniques due to their easy implementation and high efficiency. Their roots can be traced back to the first random walk problem proposed by Pearson in 1905: "*A man starts from a point O, and walks ℓ yards in a straight line; he then turns through any angle whatever, and walks another ℓ yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and r + dr from his starting point O.*" [1–4].

Using the ability of computers to generate and store large samples from multi-variate distributions, physicists and engineers have transformed the original random walk into a powerful optimization tool. Probabilistic algorithms do not require additional information on the fitness function, they simply *generate* potential candidate solutions, *select* the best, and move on.

Sharing the same random generator, yet differing with respect to the selection phase, two classes of probabilistic algorithms became more popular over the last decades: *simulated annealing* (also known as Metropolis or Hastings algorithm) [5] and *evolutionary algorithms* (EAs) [6,7]. Only the latter will be discussed in this paper.

EAs are assessed based on local quantities, such as success probability and progress rate, respectively, on global measures, like expected convergence time. Performance depends on both the fitness landscape, and on the particular probabilistic scheme (leading to a probability distribution) involved in the generating-selection mechanism. A popular test problem consists in minimizing the quadratic *SPHERE* function (To avoid confusion, we use uppercase for the fitness function, and lowercase for the uniform distribution in/on the sphere.), with optimum in the origin.

$$\mathcal{F} : \Re^n \to \Re \qquad \mathcal{F}(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^2. \tag{1}$$

An elitist (that is, keeping always the best solution found so far), one-individual, mutation+selection EA is depicted below (Algorithm 1).

Algorithm 1 An elitist, one-individual, mutation+selection EA.

Set $t = 0$ and the initial point of the algorithm, x^0
Repeat
   $t := t + 1$
   Mutation: generate a new point $x \in \Re^n$ according to a multi-variate distribution
   Selection:if $\mathcal{F}(\mathbf{x}) < \mathcal{F}(\mathbf{x}^{t-1})$ then $\mathbf{x}^t := \mathbf{x}$
        else $x^t := x^{t-1}$
Until $t = t_{\max}$, for some fixed $t_{\max}$

The region of success of Algorithm 1 is

$$R_x^S = \{y \in \Re^n | \mathcal{F}(y) < \mathcal{F}(x)\}. \tag{2}$$

A rigorous description of the long-term behavior of EAs involves renewal processes, drift analysis, order statistics, martingales, or other stochastic processes [7–13]. However, the basic structure is that of a Markov chain, as the algorithm's state at the next iteration depends only on its current state. Difficulties occur when, even for simple fitness functions, SPHERE included, the actual position of the algorithm affects significantly the local behavior, such that the process lacks homogeneity; so begins the search for powerful mathematical tools, able to describe the transition kernel, which encapsulates the local probabilistic structure of the algorithm [6,7].

For any fitness function $\mathcal{F}$ and fixed point $x \in \Re^n$, assumed as current state, the transition kernel provides the probability of the algorithm to be in set $A \subset \Re^n$ at the next step. Even if the mutation distribution has a probability density function (pdf), discontinuity occurs due to the disruptive effect of elitist selection. To make that clear, let us denote the mutation random variable by $\mathbf{Y}$, its pdf by f, and cumulative distribution function (cdf) by F. The singular (Dirac) distribution, that loads only one point, is denoted δ. Index x designates conditioning by the current state. The finite time behavior of the algorithm is inscribed in the random variable $\mathbf{Z}_x$ and the next state of the algorithm, provided the current state, is x.

$$\mathbf{Z}_x(\omega) = \begin{cases} \mathbf{Y}_x(\omega), & \mathbf{Y}_x(\omega) \in R_x^S \\ x, & \mathbf{Y}_x(\omega) \in \Re^n \setminus R_x^S \end{cases}. \tag{3}$$

while the (Markov) *transition kernel* carries the local transition probabilities

$$P_x(A) = \int_{A \cap R_x^S} f(y)dy + \left[1 - \int_{A \cap R_x^S} f(y)dy\right] \cdot \delta_x(A). \tag{4}$$

As the mutation distribution is entirely responsible for the evolution of the algorithm, let us take a look at the possible candidates, the multi-variate distributions.

Let $\mathbf{x} = (\mathbf{x_1}, \dots, \mathbf{x_n})$ denote some n-dimensional random variable, and the euclidean norm in $\Re^n$ be $||x|| = (x'x)^{1/2} = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$. The n-dimensional sphere of radius 1, its surface and volume are given by [14]

$$S_n = \{x \in \Re^n \mid ||x|| \le 1\}, \quad \delta S_n = \{x \in \Re^n \mid ||x|| = 1\}, \quad V_n = \frac{2\pi^{\frac{n}{2}}}{n\Gamma_{\frac{n}{2}}}. \tag{5}$$

We use 'bold-face' for (single, or multi-variate) random variables, and 'normal-face' for real numbers (or vectors). When partitioning the n dimensions into two sets $\{1, \dots, m\}$

and $\{m+1,\ldots,n\}$ with $1 \leq m < 1$, we use the compact notation $x = (x_{(1)}, x_{(2)}) = ((x_1,\ldots,x_m),(x_{m+1},\ldots,x_n))$, for either vectors or random variables. Unless otherwise stated, $\mathbf{Beta}_{a,b}$ denotes the beta random variable with support $(0,1)$ and parameters a, b, while β and Γ stand for the corresponding coefficients. The (one-dimensional) uniform random variable with support (a,b) is denoted $\mathbf{U}_{(a,b)}$. The sign $\stackrel{d}{=}$ denotes two random variables with identical cdf.

The class of spherical distributions can be defined in a number of equivalent ways, two of which are depicted below ([15], pp. 30, 35) (See the excellent monograph of Fang et al. [15] for an exhaustive introduction to spherical distributions.):

Definition 1.

- *An n-dimensional random variable* $\mathbf{x}$ *is said to have spherically symmetric distribution* (or simply spherical distribution) *if*

$$\mathbf{x} \stackrel{d}{=} \mathbf{r} \cdot \mathbf{u^n} \tag{6}$$

for some one-dimensional random variable (radius) $\mathbf{r}$*, and the uniform distribution on the unit sphere* $\mathbf{u^n}$*. Moreover,* $\mathbf{r}$ *and* $\mathbf{u^n}$ *are independent, and also*

$$\mathbf{r} = ||\mathbf{x}|| \geq 0, \qquad \mathbf{u^n} \stackrel{d}{=} \frac{\mathbf{x}}{||\mathbf{x}||}. \tag{7}$$

- *If the spherical distribution has pdf* g*, then* g *satisfies* $g(x) = g(||x||)$*, and there is a special connection between* g *and* f*, the pdf of* $\mathbf{r}$*, namely,*

$$f(r) = \frac{2\pi^{n/2}}{\Gamma_{\frac{n}{2}}} r^{n-1} g(r^2). \tag{8}$$

Three spherical distributions are of particular interest in our analysis:

- the uniform distribution *on* the unit sphere, with support δS_n, denoted $\mathbf{u^n}$;
- the uniform distribution *in (inside)* the unit sphere, with support S_n, denoted simply UNIFORM in this paper; and
- the standard normal distribution, denoted $\mathbf{N(0, I_n)}$ or simply NORMAL.

A comparison of the previous distributions can be performed from many angles. NORMAL was the first discovered, applied, and thoroughly analyzed in statistics, as being one of the only spherical distributions with independent and identically distributed marginals. By contrast, the components of uniform distributions on/in the sphere are not independent, neither uniform (However, the conditional marginals are uniform, see Theorem 5). Recently, the scientific interest shifted to uniform multi-variate distributions, following the increasing application of directional statistics to earth sciences and quantum mechanics [16,17], and also the application of Dirichlet distribution (which lies at the basis of spherical analysis) to Bayesian inference, involved in medicine, genetics, and text-mining [18].

From the computer science point of view, uniform and normal distributions share an entangled history, in at least two areas: random number generators and probabilistic optimization algorithms. With respect to the first area, early approaches to sampling from the uniform distribution on sphere were actually using multi-normal random generators to produce a sample x, which was further divided by $||x||$, following Equation (7). Nowadays, the situation changed, with the appearance of a new class of algorithms which circumvent the usage of the normal generator by using properties of marginal uniform distributions on/in spheres [17,19]. The comparison of mean computation times demonstrates that the uniform sampling method outperforms the normal generator for dimensions up to $n = 7$ [20].

Concerning global optimization, the probabilistic algorithms based on the two distribution types evolved at the same time, although with few overlaps. In the theory and practice

of real space (continuous) EA-*evolution strategy* being their most popular, the representative-normal distribution has played, from the beginning, the central role. Therefore, there is a great amount of literature stressing out the advantages of this distribution [6,21,22].

Occasionally, the supremacy of the normal operator has been challenged by theoretical studies that proposed different mutation distributions, such as uniform in the cube (which is non-spherical) [10], uniform *on* the sphere [7,23], uniform *in* the sphere [9], or even the Cauchy distribution [24]. An attempt to solve the problem globally, by considering the whole class of *spherical (isotropic)* distributions, was made in [25,26]. These approaches yielded only limited results (valid either for small space dimension *n*, or for $n \to \infty$), or not so tight lower/upper bounds for the expected progress of the algorithm.

The study of EAs with uniform distribution in the sphere recently culminated with two systematic studies, one for RIDGE [27], the other for SPHERE [28], comparable to classical theory of evolution strategies [6,29]. Under a carefully constructed normalization of mutation parameters (equalizing the expectations of normal and uniform multi-variates as $n \to \infty$), those studies demonstrate the same behavior for the respective EA variants. Intuitively, the explanation is that for large dimensions, both normal and uniform distributions concentrate *on* the surface of the sphere. The present paper differs from the previous analyses in the way that it does not apply any normalization of parameters. As a consequence, the results are different from those in [28] and an actual comparison between the two algorithms can be achieved.

Section 2 discusses the general framework of spherical multi-variate distributions, with special focus on uniform and normal. Then, two algorithms, one with uniform mutation and the other with normal mutation, are compared on the SPHERE fitness function with respect to their local performance in Section 3.

2. Materials and Methods. Spherical Distributions

In light of Definition 1, the spherical distributions are very much alike. They all exhibit stochastic representation (6), that is, each can be generated as a product of two independent distributions, the *n*-dimensional uniform on the sphere $\mathbf{u^n}$ and some scalar, positive random variable $\mathbf{r}$. As the distribution of $\mathbf{r}$ makes the whole difference, we point out the form of this random variable in the three cases of interest.

- $\mathbf{u^n}$-$\mathbf{r}$ is obviously the Dirac distribution in 1:

$$\delta_1(x) = 1, \quad x = 1. \tag{9}$$

- NORMAL-$\mathbf{r}$ is the χ distribution with *n* degrees of freedom, with pdf ([7], p. 20):

$$f(x) = \frac{1}{2^{\frac{n}{2}-1}\Gamma_{\frac{n}{2}}} e^{-\frac{x^2}{2}} x^{n-1}, \quad x \in (0, \infty). \tag{10}$$

- UNIFORM-$\mathbf{r}$ is distributed $\mathbf{Beta}_{n,1}$, with pdf ([15], p. 75):

$$f(x) = nx^{n-1}, \quad x \in (0, 1). \tag{11}$$

Using as primary source the monograph [15], we next discuss in more detail the stochastic properties of the UNIFORM and NORMAL multi-variates, the two candidates for the mutation operator of the algorithm.

2.1. Uniform in the Sphere

The local analysis of the EA is based on two particular marginal distributions: the first component $\mathbf{x_1}$, and the joint marginal of the remaining $n - 1$ components, $\mathbf{x_{(2)}}$. As already pointed out, the marginals of UNIFORM are not independent random variables, and we shall see that neither are they uniform. A general formula for the marginal density is provided in [15] (p. 75):

Theorem 1. *If* $\mathbf{x} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)})$ *is uniformly distributed in the unit sphere, with* $\mathbf{x}_{(1)}$ *of dimension* k, $1 \leq k < n$, *then the marginal density of* $\mathbf{x}_{(1)}$ *is*

$$f(x_{(1)}) = \frac{\pi^{-\frac{k}{2}} \Gamma_{\frac{n+2}{2}}}{\Gamma_{\frac{n-k+2}{2}}} \left(1 - ||x_{(1)}||\right)^{\frac{n-k}{2}}, \quad ||x_{(1)}||^2 < 1. \tag{12}$$

Corollary 1. *The pdf of the first component of UNIFORM is*

$$f(x) = \frac{1}{\beta_{\frac{n+1}{2}, \frac{1}{2}}} \left(1 - x^2\right)^{\frac{n-1}{2}}, \quad x \in (-1, 1). \tag{13}$$

Using the symmetry with respect to the origin and substituting $x^2 = t$ in function (13), we obtain an interesting result, previously unreported in spherical distributions literature.

Corollary 2. *The square of the first component of UNIFORM is* **Beta**$_{\frac{n+1}{2}, \frac{1}{2}}$*, with pdf*

$$f(x) = \frac{1}{\beta_{\frac{n+1}{2}, \frac{1}{2}}} (1 - x)^{\frac{n-1}{2}} x^{-\frac{1}{2}}, \quad x \in (0, 1). \tag{14}$$

The density of the last $n - 1$ components can be derived also from Theorem 1.

Corollary 3. *The joint pdf of the last* $n - 1$ *components of UNIFORM is*

$$f(x_{(2)}) = n\pi^{-\frac{n}{2}} \Gamma_{\frac{n}{2}} \left(1 - ||x_{(2)}||^2\right)^{\frac{1}{2}}, \quad ||x_{(2)}||^2 < 1. \tag{15}$$

As function of several variables, formula (15) might not look very appealing; however, a basic result from spherical distribution theory transforms the corresponding multiple integral into a scalar one ([15], p. 23).

Theorem 2 (Dimension reduction)**.**

$$\int f\left(\sum_{i=1}^{m} x_i^2\right) dx_1, \ldots, dx_m = \frac{\pi^{\frac{m}{2}}}{\Gamma_{\frac{m}{2}}} \int_0^{\infty} y^{\frac{m}{2}-1} f(y) dy. \tag{16}$$

One can see now that, if $\mathbf{x}$ is uniformly distributed in the unit sphere, the sum of squares of the last $n - 1$ components is **Beta** distributed.

Corollary 4. *Let* $\mathbf{x} = (\mathbf{x_1}, \mathbf{x}_{(2)})$ *be UNIFORM. Then, the one-dimensional random variable* $||x_{(2)}||^2$ *is* **Beta**$_{\frac{n-1}{2}, \frac{3}{2}}$*, with pdf*

$$f(x) = \frac{1}{\beta_{\frac{n-1}{2}, \frac{3}{2}}} (1 - x)^{\frac{n-3}{2}} x^{\frac{1}{2}}, \quad x \in (0, 1). \tag{17}$$

Proof. Apply transformation (16) to (15), $m = n - 1$ and $f(y) = (1 - y)^{(n-1)/2}$. □

As the components of the uniform distribution on/in the sphere are not independent, a better understanding of the nature of such distributions is provided by conditioning one component with respect to the others. In case of uniform distribution *on* the sphere, the work in [15] (p. 74) states that all the conditional marginals are also uniform *on* the sphere.

We shall see that a similar characterization holds true for the uniform *in* the sphere. This result is not presented in [15] and, to the best of our knowledge, in no other reference on spherical distributions. Therefore, an additional theorem is needed ([30], p. 375).

Theorem 3. *Let* $\mathbf{x} \stackrel{d}{=} \mathbf{r} \cdot \mathbf{u^n}$ *be a spherical distribution and* $\mathbf{x} = (\mathbf{x_{(1)}}, \mathbf{x_{(2)}})$, *where* $\mathbf{x_{(1)}}$ *is* m-*dimensional,* $1 \leq m < n$. *Then the conditional distribution of* $\mathbf{x_{(1)}}$ *given* $\mathbf{x_{(2)}} = h$ *with* $||h|| = a$ *is given by*

$$\left(\mathbf{x_{(1)}} \mid \mathbf{x_{(2)}} = h\right) \stackrel{d}{=} \mathbf{r}_{a^2} \cdot \mathbf{u^m}. \tag{18}$$

For each $a \geq 0$, $\mathbf{r}_{a^2}$ *and* $\mathbf{u^m}$ *are independent, and the cdf of* $\mathbf{r}_{a^2}$ *is given by*

$$prob(\mathbf{r}_{a^2} \leq t) = \frac{\int_a^{\sqrt{t^2+a^2}} (r^2 - a^2)^{m/2-1} \, r^{-(n-2)} dF(r)}{\int_a^{\infty} (r^2 - a^2)^{m/2-1} \, r^{-(n-2)} dF(r)}, \tag{19}$$

for $t \geq 0$, $a > 0$ *and* $F(a) < 1$, F *being the cdf of* $\mathbf{r}$.

We prove now the result on conditional marginals of the uniform distribution *in* the unit sphere. As conditioning the first component with respect to all others is the most relevant for EA analysis, this particular case is stressed out.

First, an old result from probability theory is needed, similar to the convolution operation, but for the *product* of two independent random variables [31].

Theorem 4 (Mellin's formula)**.** *Let* $\mathbf{y}$ *and* $\mathbf{z}$ *be two independent, non-negative random variables, with densities* g *and* h. *Then,* $\mathbf{x} = \mathbf{y} \cdot \mathbf{z}$ *has pdf*

$$f(x) = \int_0^{\infty} \frac{1}{z} \, g\left(\frac{x}{z}\right) h(z) dz, \quad x \in (0, \infty). \tag{20}$$

Note that Mellin's formula still holds, if only one of the random variables is continuous, the other being discrete, see, e.g., in [15] (p. 41).

Theorem 5.

- *Let* $\mathbf{x} = (\mathbf{x_{(1)}}, \mathbf{x_{(2)}})$ *be UNIFORM, where* $\mathbf{x_{(1)}}$ *is* m-*dimensional,* $1 \leq m < n$. *The conditional distribution of* $\mathbf{x_{(1)}}$ *given* $\mathbf{x_{(2)}} = a$ *is UNIFORM in the* m *dimensional sphere with radius* $\left(1 - ||a||^2\right)^{1/2}$.
- *If* $m = 1$ *and* $x_{(2)} = h$ *is a point in* S_{n-1} *with* $||h|| = a$, *the conditional distribution of* $\mathbf{x_1}$ *given* $\mathbf{x_{(2)}} = h$ *is*

$$\left(\mathbf{x_1} \mid \mathbf{x_{(2)}} = h\right) \stackrel{d}{=} \mathbf{U}_{\left(-\sqrt{1-a^2}, \sqrt{1-a^2}\right)}. \tag{21}$$

Proof. We begin with the last part, case $m = 1$.

Equation (18) gives the conditional first component as a product of two independent random variables, the second being the one-dimensional UNIFORM-the discrete random variable that loads -1 and 1 with equal probability, $1/2$.

The cdf of the first random variable is given by (19), as a fraction of two integrals, both with respect to F, the cdf of $\mathbf{r}$. In case of UNIFORM, $\mathbf{r}$ is $\mathbf{Beta}_{n,1}$ given by (11), thus $dF(r) = f(r)dr = nr^{n-1}dr$.

If $t \geq \sqrt{1 - a^2}$, the upper and lower integrals in (19) are equal, so the probability is 1. If $0 \leq t < \sqrt{1 - a^2}$, the upper integral is

$$n \int_a^{\sqrt{t^2+a^2}} r(r^2 - a^2)^{-1/2} dr = nt,$$

while the lower one is

$$n \int_a^1 r(r^2 - a^2)^{-1/2} dr = n\sqrt{1 - a^2}.$$

Thus, for any fixed $a < 1$, the cdf of the conditional radius is

$$prob(\mathbf{r}_{a^2} \leq t) = \frac{t}{\sqrt{1 - a^2}}, \quad t \in (0, \sqrt{1 - a^2}),$$

while the corresponding pdf is

$$\mathbf{r}_{a^2} \stackrel{d}{=} \mathbf{U}_{(0, \sqrt{1-a^2})}.$$

Back to the application of Theorem 3, Equation (18). The conditional first component of UNIFORM is a product of two independent random variables: one continuous, the other discrete. This is the easy version of Mellin's formula, and the result is the continuous uniform random variable with support $(-\sqrt{1 - a^2}, \sqrt{1 - a^2})$ from Equation (21).

As for a larger dimension, $m > 1$, the cdf in (19) becomes

$$prob(\mathbf{r}_{a^2} \leq t) = \frac{t^m}{(1 - a^2)^{\frac{m}{2}}}, \quad t \in \left(0, (1 - a^2)^{\frac{m}{2}}\right),$$

with corresponding pdf

$$f(t) = \frac{mt^{m-1}}{(1 - a^2)^{\frac{m}{2}}}, \quad t \in \left(0, (1 - a^2)^{\frac{m}{2}}\right),$$

which is $\mathbf{Beta}_{m,1}$, yet with reduced support.

Summing up, Equation (18) provides the conditional marginal as a product of the uniform in the unit sphere and a reduced $\mathbf{Beta}_{m,1}$. According to representation (11), the result is UNIFORM, in m dimensions, with the center being the origin and reduced radius $r = (1 - a^2)^{m/2}$. $\quad\square$

2.2. Normal

As we did with UNIFORM, we denote the first component of the standard normal multi-variate distribution by $\mathbf{x}$, and the remaining $n - 1$ components by $\mathbf{x}_{(2)}$. Due to independence of the marginals, one can write a compact equivalent of Propositions 1 and 3, see, e.g., in [6] (p. 54).

Proposition 1. *Let* $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_{(2)}})$ *be NORMAL.*

- *The pdf of the first component,* $\mathbf{x_1}$*, is*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \Re. \tag{22}$$

- *The joint pdf of the last* $n - 1$ *components,* $\mathbf{x_{(2)}}$*, is*

$$f(x_{(2)}) = \frac{1}{\left(\sqrt{2\pi}\right)^{n-1}} e^{-\frac{\|x_{(2)}\|^2}{2}}, \quad x_{(2)} \in \Re^{n-1}. \tag{23}$$

Due to sphericity of the joint $n - 1$ components, one obtains again a compact form for the sum of squares.

Corollary 5. *The one-dimensional random variable* $\|x_{(2)}\|^2$ *is* χ^2 *with* $n - 1$ *degrees of freedom, with pdf*

$$f(x) = \frac{2^{-\frac{n-1}{2}}}{\Gamma_{\frac{n-1}{2}}} x^{\frac{n-3}{2}} e^{-\frac{x}{2}}, \quad x \in (0, \infty). \tag{24}$$

3. Results

We restrict the study to the case P (current algorithm position) nearby O (optimum of SPHERE) and analyze the local performance of two EAs, one with uniform, the other with normal mutation, in terms of success probability and expected progress. Namely, we assume $R = |OP| \leq 1/2$, such that success region R_P^S is the sphere with center O and radius R, Figure 1.
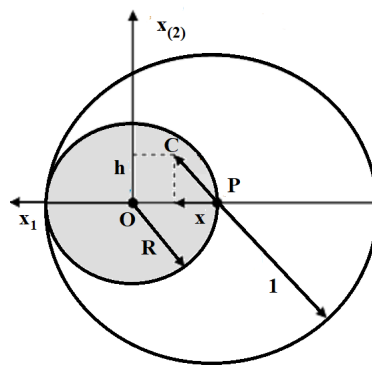

Figure 1. Success region of Algorithm 1 with uniform mutation on SPHERE.

Re-set P as the origin of the coordinate system, and measure the algorithm's progress in the positive direction of the first axis-write x for x_1, h for $x_{(2)}$ and u for $||h||^2$. The success probability and the expected progress are provided by the first term of transition kernel (4), respectively, by the upper part of random variable (3). They obey to the uniform continuous mutation distribution with pdf $f(x) = 1/V_n$ and compact support, the sphere with center P and radius 1.

The calculus of success probability and expected progress resides in integrating the random variable (3) over R_P^S. For UNIFORM mutation this calculus is analytically tractable. For NORMAL mutation, the analytic integration is impossible, see in [6] (p. 56) and Theorem 8, but the theory of incomplete Gamma functions makes the comparison tractable.

Note that, if the success probability bears only one possible definition (the volume of success region), the situation is different with respect to the expected progress. As the random variable $\mathbf{Z}_x$ from (3) characterizes the local behavior of the algorithm, one would normally associate the expected progress to the expected value of this random variable. However, $\mathbf{Z}_x$ is n-dimensional, and such is $E(\mathbf{Z}_x)$, so there is a need to mediate somehow among the n components.

One could consider only the first component of the expected value, the one pointing towards the optimum, which has been applied on a different fitness landscape, the inclined plane [21]. (Yet in another landscape, the RIDGE, it is customary to consider the progress along the perpendicular component h, see in [6,27] for an inventory of fitness functions used in EA testing, the reader is referred to the work in [32].) For UNIFORM mutation, a simplified version of the expected progress may be defined as the centroid of the corresponding success region [9,10]. However, a more traditional view is followed here ([6], p. 54):

$$\mathbf{progress} = R - \sqrt{(R-x)^2 + u}. \tag{25}$$

This corresponds to the difference in distance $|OP| - |OC|$, provided $\mathbf{C}$ is a random point generated by mutation, Figure 1.

3.1. Uniform Mutation

If the UNIFORM mutation in the unit sphere with center P is applied, one cannot use for integration the ensemble of Propositions 1 and 3, as the marginals of UNIFORM are not independent. Instead, one should use the conditional first component from Theorem 5,

together with the joint $n - 1$ dimensional distribution from Proposition 3. The integration region is

$$\begin{cases} u = ||h||^2 \in (0, R^2) \\ x \in \left(R - \sqrt{R^2 - u}, R + \sqrt{R^2 - u} \right) \end{cases}.$$

Theorem 6. *Let an EA with UNIFORM mutation minimizing the SPHERE be situated at current distance R from the origin, $R \in (0, 1/2)$. The success probability is*

$$prob^U = R^n. \tag{26}$$

Proof. The use of Equations (4), (15), (16) and (21) yields

$$prob^U = \int_{S_n^O} \mathbf{1}_{S_n^P} dx$$

$$= n\pi^{-\frac{n}{2}} \Gamma_{\frac{n}{2}} \int_{h \in S_{n-1}^O} \left(1 - ||h||^2 \right)^{\frac{1}{2}} \times \frac{1}{2\sqrt{1 - ||h||^2}} \int_{R - \sqrt{R^2 - ||h||^2}}^{R + \sqrt{R^2 - ||h||^2}} 1 \, dx \, dh$$

$$= \frac{n\pi^{\frac{n-1}{2} - \frac{n}{2}} \Gamma_{\frac{n}{2}}}{\Gamma_{\frac{n-1}{2}}} \int_0^{R^2} y^{\frac{n-1}{2} - 1} \sqrt{R^2 - y} \, dy = \frac{n \Gamma_{\frac{n}{2}}}{\sqrt{\pi} \Gamma_{\frac{n-1}{2}}} \int_0^{R^2} y^{\frac{n-3}{2}} \sqrt{R^2 - y} \, dy = R^n. \tag{27}$$

□

Theorem 7. *Let an EA with UNIFORM mutation minimizing the SPHERE be situated at current distance R from the origin, $R \in (0, 1/2)$. The expected progress is*

$$\phi^U = \frac{R^{n+1}}{n + 1} \tag{28}$$

Proof. Following the proof of Theorem 6 and inserting factor (25), one gets

$$\phi^U = \frac{1}{2} \frac{n \Gamma_{\frac{n}{2}}}{\sqrt{\pi} \Gamma_{\frac{n-1}{2}}} \times \int_{u=0}^{R^2} u^{\frac{n-3}{2}} \int_{R - \sqrt{R^2 - u}}^{R + \sqrt{R^2 - u}} \left(R - \sqrt{(R - x)^2 + u} \right) dx \, du$$

$$= C \int_{u=0}^{R^2} u^{\frac{n-3}{2}} \int_{R - \sqrt{R^2 - u}}^{R + \sqrt{R^2 - u}} R \, dx \, du - C \int_{u=0}^{R^2} u^{\frac{n-3}{2}} \int_{R - \sqrt{R^2 - u}}^{R + \sqrt{R^2 - u}} \sqrt{(R - x)^2 + u} \, dx \, du$$

$$= C \, I_1 - C \, I_2.$$

We treat separately I_1 and I_2. I_1 is simply (27), multiplied by the constant R, thus

$$C \, I_1 = R^{n+1}. \tag{29}$$

For I_2 one can apply formula ([33], p. 13)

$$\int (x^2 + a)^{\frac{1}{2}} dx = \frac{x}{2}(x^2 + a)^{\frac{1}{2}} + \frac{a}{2} \ln \left(x + (x^2 + a)^{\frac{1}{2}} \right).$$

in order to get

$$I_2 = \int_{u=0}^{R^2} u^{\frac{n-3}{2}} R \sqrt{R^2 - u} \, du + \int_{u=0}^{R^2} u^{\frac{n-3}{2}} \frac{u}{2} \ln \frac{R + \sqrt{R^2 - u}}{R - \sqrt{R^2 - u}} \, du = I_3 + I_4. \tag{30}$$

Again, I_3 is the integral (27), multiplied by $R/2$, thus

$$C \, I_3 = \frac{R^{n+1}}{2}. \tag{31}$$

The substitution $y = u/R^2$ on I_4 provides

$$I_4 = \frac{R^{n+1}}{2} \int_0^1 y^{\frac{n-1}{2}} \ln \frac{1 + \sqrt{1-y}}{1 - \sqrt{1-y}} \, dy$$

while partial integration gives

$$I_4 = \frac{R^{n+1}}{n+1} \int_0^1 y^{\frac{n+1}{2}} \cdot \frac{1}{y\sqrt{1-y}} \, dy = \frac{R^{n+1}}{n+1} \beta_{\frac{n+1}{2}, \frac{1}{2}}.$$

Bringing in the constant C, one gets

$$C\, I_4 = \frac{1}{2} \frac{n\Gamma_{\frac{n}{2}}}{\sqrt{\pi}\Gamma_{\frac{n-1}{2}}} \cdot \frac{R^{n+1}}{n+1} \beta_{\frac{n+1}{2}, \frac{1}{2}} = \frac{1}{2} \frac{n-1}{n+1} R^{n+1}. \tag{32}$$

Summing up (29)–(32) one gets the desired result. $\square$

The results from Theorems 6 and 7 are also presented in [28], yet with different proofs. Equations (26) and (28) point out a remarkable property of the EA with UNIFORM mutation on the SPHERE.

Corollary 6. *In the conditions of Theorems 6 and 7, the success probability is the derivative of the expected progress.*

3.2. Normal Mutation

Setting $\sigma = 1$ and avoiding the transformation $\sigma^* = \sigma n/R$, one obtains the success probability and the expected progress for the EA with standard normal mutation following closely the proof in ([6], pp. 54–56). The incomplete Gamma function is ([33], p. 260):

$$P(a, x) = \frac{1}{\Gamma_a} \int_0^x e^{-t} t^{a-1} dt. \tag{33}$$

The following expressions are not restricted to the case of algorithm nearby optimum, due to the unbounded support of the normal distribution. Unfortunately, integration is impossible.

Theorem 8. *Let an EA with NORMAL mutation minimizing the SPHERE be situated at current distance R from the origin.*

- *The success probability is*

$$prob^N = \frac{1}{\sqrt{2\pi}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} P\left(\frac{n-1}{2}, Rx - \frac{x^2}{2}\right) dx. \tag{34}$$

- *The expected progress is*

$$\phi^N = \frac{2^{-\frac{n-1}{2}}}{\sqrt{2\pi}\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{2R} \int_{u=0}^{2Rx-x^2} e^{-\frac{x^2}{2}} u^{\frac{n-1}{2}-1} e^{-\frac{u}{2}} \, progress \, du dx. \tag{35}$$

Proof.

$$
prob^N = \frac{2^{-\frac{n-1}{2}}}{\sqrt{2\pi}\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{2R} \int_{u=0}^{2Rx-x^2} e^{-\frac{x^2}{2}} u^{\frac{n-1}{2}-1} e^{-\frac{u}{2}} du\,dx
$$

$$
= \frac{2^{-\frac{n-1}{2}}}{\sqrt{2\pi}\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} \left(\int_{u=0}^{2Rx-x^2} u^{\frac{n-1}{2}-1} e^{-\frac{u}{2}} du \right) dx
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} \left(\frac{1}{\Gamma_{\frac{n-1}{2}}} \int_{t=0}^{Rx-\frac{x^2}{2}} t^{\frac{n-1}{2}-1} e^{-t} dt \right) dx
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} P\left(\frac{n-1}{2}, Rx - \frac{x^2}{2} \right) dx. \tag{36}
$$

The same calculus applies for the expected progress, with the addition of the factor corresponding to the one dimensional progress along the x axis, Equation (25). □

3.3. Comparison

Due to the analytic intractability of integral representations (34) and (35), a theoretical comparison between two variants of Algorithm 1-one with UNIFORM, the other with NORMAL mutation-must resort to inequalities. Therefore, a deeper insight into the prolific theory of Euler and hypergeometric functions is required. We start with an upper bound for the incomplete Gamma function (33) ([34], p. 1213).

Proposition 2. *The following inequality holds*

$$
P(a, x) \leq \frac{x^a}{a(a+1)\,\Gamma_a} \left(1 + ae^{-x} \right). \tag{37}
$$

More results are gathered from in [35], [36] (p. 240), [37] (pp. 890, 894), [38] (pp. 53, 57).

Proposition 3 (Hypergeometric functions).

• *For any real set of parameters a, b, a_i, b_i and any real number x, define*

$$
{}_1F_1\left(a, b \mid x\right) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{x^k}{k!} \tag{38}
$$

$$
{}_2F_2\left(a_1, a_2; b_1, b_2 \mid x\right) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k}{(b_1)_k (b_2)_k} \frac{x^k}{k!} \tag{39}
$$

where $(a)_k = a(a+1)\ldots(a+k-1) = \Gamma_{a+k}/\Gamma_a$ is the Pochhammer symbol, with

$$
(a)_{2k} = \left(\frac{a}{2}\right)_k \left(\frac{a+1}{2}\right)_k 2^{2k}. \tag{40}
$$

• *If $A = (a_1, \ldots, a_q)$ and $B = (b_1, \ldots, b_q)$, we write $B \prec^W A$, if*

$$
0 < a_1 \leq \ldots \leq a_q, \quad 0 < b_1 \leq \ldots \leq b_q
$$

$$
\sum_{i=1}^{k} a_i \leq \sum_{i=1}^{k} b_i, \quad k = 1, \ldots, q.
$$

- If $B \prec^W A$, the following inequality holds:

$$_pF_p\left(A, B \mid x\right) \leq 1 - \theta + \theta e^x, \tag{41}$$

$$\theta = \begin{cases} \frac{a}{b}, & p = 1, \ (A, B) = (a, b) \\ \frac{a_1 a_2}{b_1 b_2}, & p = 2, \ (A, B) = (a_1, a_2; b_1, b_2) \end{cases}.$$

We can prove now the main result stating that, for an EA acting on the SPHERE with current position at maximal range $1/2$ from the origin, the UNIFORM mutation provides a larger success probability than the NORMAL mutation, for the arbitrary dimension n. We denote by $n!!$ the double factorial (semi-factorial), that is, the product of all integers from 1 to n of same parity with n.

Theorem 9. *Let an EA minimizing the SPHERE be situated at current distance R from the origin, such that $R \leq 1/2$. For any $n \geq 3$, the following holds:*

$$\frac{prob^N}{prob^U} < \frac{1}{n!!}. \tag{42}$$

Proof. Apply inequality (37) to Equation (34)

$$prob^N \leq \frac{1}{\sqrt{2\pi}\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} \frac{4\left(Rx - \frac{x^2}{2}\right)^{\frac{n-1}{2}}}{n^2 - 1} \left(1 + \frac{n-1}{2}e^{\frac{x^2}{2} - Rx}\right) dx$$

$$= \frac{4}{\sqrt{2\pi}(n^2 - 1)\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{2R} e^{-\frac{x^2}{2}} \left(Rx - \frac{x^2}{2}\right)^{\frac{n-1}{2}} dx$$

$$+ \frac{2}{\sqrt{2\pi}(n+1)} \int_{x=0}^{2R} e^{-Rx}\left(Rx - \frac{x^2}{2}\right)^{\frac{n-1}{2}} dx$$

$$= \frac{R^n 2^{\frac{n+5}{2}}}{\sqrt{2\pi}(n^2 - 1)\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{1} e^{-2R^2 t^2} t^{\frac{n-1}{2}} (1 - t)^{\frac{n-1}{2}} dt \tag{43}$$

$$+ \frac{R^n 2^{\frac{n+3}{2}}}{\sqrt{2\pi}(n+1)\Gamma_{\frac{n-1}{2}}} \int_{x=0}^{1} e^{-2R^2 t} t^{\frac{n-1}{2}} (1 - t)^{\frac{n-1}{2}} dt.$$

Using the series expansion of the exponential, the second integral in (43) becomes a hypergeometric function of type (38). The interchange of the integral and the sum is justified by the absolute convergence of the series.

$$\int_{x=0}^{1} e^{-2R^2 t} t^{\frac{n-1}{2}} (1 - t)^{\frac{n-1}{2}} dt = \int_{x=0}^{1} t^{\frac{n-1}{2}} (1 - t)^{\frac{n-1}{2}} \sum_{k=0}^{\infty} \frac{(-2R^2 t)^k}{k!} dt$$

$$= \sum_{k=0}^{\infty} \frac{(-2R^2)^k}{k!} \int_{x=0}^{1} t^{\frac{n-1}{2}+k} (1 - t)^{\frac{n-1}{2}} dt = \sum_{k=0}^{\infty} \frac{(-2R^2)^k}{k!} \frac{\Gamma_{\frac{n+1}{2}+k}\Gamma_{\frac{n+1}{2}}}{\Gamma_{n+1+k}}$$

$$= \beta_{\frac{n+1}{2}, \frac{n+1}{2}} \sum_{k=0}^{\infty} \frac{(-2R^2)^k \left(\frac{n+1}{2}\right)_k}{k! (n+1)_k} = 2^{-n}\beta_{\frac{n+1}{2}, \frac{1}{2}} \, _1F_1\left(\frac{n+1}{2}, n+1 \mid -2R^2\right). \tag{44}$$

Identity (40) reduces the first integral in (43) to a hypergeometric function (39).

$$
\int_{x=0}^{1} e^{-2R^2 t^2} t^{\frac{n-1}{2}} (1-t)^{\frac{n-1}{2}} dt = \int_{x=0}^{1} t^{\frac{n-1}{2}} (1-t)^{\frac{n-1}{2}} \sum_{k=0}^{\infty} \frac{\left(-2R^2 t^2\right)^k}{k!} dt
$$

$$
= \sum_{k=0}^{\infty} \frac{\left(-2R^2\right)^k}{k!} \int_{x=0}^{1} t^{\frac{n-1}{2}+2k} (1-t)^{\frac{n-1}{2}} dt = \sum_{k=0}^{\infty} \frac{\left(-2R^2\right)^k}{k!} \frac{\Gamma_{\frac{n+1}{2}+k} \Gamma_{\frac{n+1}{2}}}{\Gamma_{n+1+k}}
$$

$$
= \beta_{\frac{n+1}{2},\frac{n+1}{2}} \sum_{k=0}^{\infty} \frac{\left(-2R^2\right)^k}{k!} \frac{\left(\frac{n+1}{2}\right)_{2k}}{(n+1)_{2k}} = \beta_{\frac{n+1}{2},\frac{n+1}{2}} \sum_{k=0}^{\infty} \frac{\left(-2R^2\right)^k}{k!} \frac{\left(\frac{n+1}{4}\right)_k \left(\frac{n+3}{4}\right)_k 2^{2k}}{\left(\frac{n+1}{2}\right)_k \left(\frac{n+2}{2}\right)_k 2^{2k}}
$$

$$
= 2^{-n} \beta_{\frac{n+1}{2},\frac{1}{2}}\, {}_2F_2\left(\frac{n+1}{4}, \frac{n+3}{4}; \frac{n+1}{2}, \frac{n+2}{2}\bigg| -2R^2\right). \tag{45}
$$

Summing up Equations (43)–(45), and using inequality (41) for $p = 1, 2$, one gets

$$
prob^N \le \frac{R^n}{2^{\frac{n+2}{2}}(n+1)\Gamma_{\frac{n+2}{2}}} \left[(n-1)\left(1+e^{-2R^2}\right) + 4 - \frac{n+3}{n+2}\left(1-e^{-2R^2}\right)\right]
$$

$$
< \frac{R^n}{2^{\frac{n+2}{2}}(n+1)\Gamma_{\frac{n+2}{2}}} [2(n-1)+4] = R^n \frac{1}{2^{\frac{n}{2}}\Gamma_{\frac{n+2}{2}}} = R^n \frac{1}{2^{\frac{n}{2}}\frac{n}{2}\frac{n-2}{2}\cdots} = prob^U \frac{1}{n!!}.
$$

In the last equality we have used Equation (26) and the definition of the double factorial, for n even. Obviously, for n odd the constant $\sqrt{2}/\sqrt{\pi}$ will appear at the tail of the product, yet this is a minor difference that may be neglected. □

The result for the expected progress follows now easily.

Theorem 10. *Let an EA minimizing the SPHERE be situated at current distance R from the origin, such that $R \in (0, 1/2)$. For any $n > 3$, the following holds:*

$$
\frac{\phi^N}{\phi^U} < \frac{n+1}{n!!} \approx \frac{1}{(n-2)!!}. \tag{46}
$$

Proof. The expected progress of NORMAL mutation (35) differs from success probability (34) only by the integration factor (25). As **progress** $< R$, inequality (46) is a simple consequence of Theorems 7 and 9. □

4. Discussion

Within evolutionary algorithms acting on real space, the use of normal distribution makes the implementation easier: in order to generate an n dimensional point, one simply generates n times from the normal uni-variate. Unfortunately, simplicity of the practical algorithm does not transfer to the theoretical analysis, making EA experts go long distances in order to estimate performance quantities like the success probability and the expected progress. In the end, the normal mutation only provides asymptotic formulas, valid for large n.

This paper analyzes a different mutation operator, based on the uniform multi-variate in the sphere, with dependent components. Using deeper insights into the spherical distributions theory, the local performance of the algorithm with uniform mutation was measured on the SPHERE fitness function. Close expressions for the success probability and the expected progress of the EA with uniform mutation have been derived, valid for arbitrary n. Compared to the performance of the normal operator-which, due to the intractability of integral formulas in Theorem 8, required inequalities with hypergeometric functions-, the success probability and the expected progress of the algorithm with uniform mutation are both larger, by a factor of order $n!!$.

5. Conclusions

From a broader perspective, this paper can be seen, together with the works in [27,28], as an attempt of revisiting the classical theory of continuous evolutionary algorithms. Even if practitioners in the field will continue to use the normal multi-variate as mutation distribution, we claim that the theory can benefit from the uniform distribution inside the sphere. First, as demonstrated in this paper, a particular setting of parameters (the natural choice $\rho = \sigma = 1$) provides better performance for the uniform mutation operator on the SPHERE landscape, if current position of the algorithm is nearby the optimum. However, in light of the "no free-lunch theorem for optimization" paradigm [39], one cannot expect general dominance of an algorithm over all others, irrespective of the fitness function. Rather, specific algorithms with particular operators should be analyzed separately, on different optimization landscapes. This is where the second advantage of the new uniform distribution occurs, in terms of more tractable mathematical analysis, yielding close formulas, previously not attained by normal mutation theory—see the studies of the RIDGE landscape in [28] and of the elitist evolutionary algorithm with mutation and crossover on SPHERE in [27].

A theory of continuous evolutionary algorithms could not be complete without the analysis of global behavior and adaptive mutation parameter. These cases have already been treated in [27,28]—under a normalization of mutation sphere radius which makes algorithm behave similarly to the one with normal mutation, in terms of difference and differential equations, following the works in [6,29]. This opens the way for the challenging task, previously unattempted in probabilistic optimization literature, of linking the theory of continuous evolutionary algorithms to that of differential optimization techniques such as particle swarm optimization [40] and differential evolution [41].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pearson, K. The problem of the random walk. *Nature* **1905**, *72*, 294. [CrossRef]
2. Kluyver, J.C. A local probability problem. *Nederl. Acad. Wetensch. Proc.* **1905**, *8*, 341–350.
3. Watson, G.N. *A Treatise on the Theory of Bessel Functions*; University Press: Cambridge, UK, 1995.
4. Zhou, Y. On Borwein's conjectures for planar uniform random walks. *J. Aust. Math. Soc.* **2019**, *107*, 392–411. [CrossRef]
5. Dunson, D.B.; Johndrow, J.E. The Hastings algorithm at fifty. *Biometrika* **2020**, *107*, 1–23. [CrossRef]
6. Beyer, H.-G. *The Theory of Evolution Strategies*; Springer: Heidelberg, Germany, 2001.
7. Rudolph, G. *Convergence Properties of Evolutionary Algorithms*; Kovać: Hamburg, Germany, 1997.
8. Agapie, A.; Wright, A.H. Theoretical analysis of steady state genetic algorithms. *Appl. Math.* **2014**, *59*, 509–525. [CrossRef]
9. Agapie, A.; Agapie, M.; Rudolph, G.; Zbaganu, G. Convergence of evolutionary algorithms on the *n*-dimensional continuous space. *IEEE Trans. Cybern.* **2013**, *43*, 1462–1472. [CrossRef] [PubMed]
10. Agapie, A.; Agapie, M.; Zbaganu, G. Evolutionary Algorithms for Continuous Space Optimization. *Int. J. Syst. Sci.* **2013**, *44*, 502–512. [CrossRef]
11. Agapie, A. Estimation of Distribution Algorithms on Non-Separable Problems. *Int. J. Comp. Math.* **2010**, *87*, 491–508. [CrossRef]
12. Agapie, A. Theoretical analysis of mutation-adaptive evolutionary algorithms. *Evol. Comp.* **2001**, *9*, 127–146. [CrossRef]
13. Auger, A. Convergence results for the $(1, \lambda)$-SA-ES using the theory of ϕ-irreducible Markov chains. *Theor. Comput. Sci.* **2005**, *334*, 35–69. [CrossRef]
14. Li, S. Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian J. Math. Stat.* **2011**, *4*, 66–70. [CrossRef]
15. Fang, K.-T.; Kotz, S.; Ng, K.-W. *Symmetric Multivariate and Related Distributions*; Chapman and Hall: London, UK, 1990.
16. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; Wiley: New York, NY, USA, 2000.
17. Watson, G.S. *Statistics on Spheres*; University of Arkansas Lecture Notes in the Mathematical Sciences; Wiley: New York, NY, USA, 1983.
18. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993—1022.

19. Fang, K.-T.; Yang, Z.; Kotz, S.; Ng, K.-W. Generation of multivariate distributions by vertical density representation. *Statistics* **2001**, *35*, 281–293. [CrossRef]
20. Harman, R.; Lacko, V. On decompositional algorithms for uniform sampling from n-spheres and n-balls. *J. Multivar. Anal.* **2010**, *101*, 2297—2304. [CrossRef]
21. Rechenberg, I. *Evolutionsstrategie: Optimierung Technischer Systeme Nach Prinzipiender Biologischen Evolution*; Frommann-Holzboog Verlag: Stuttgart, Germany, 1973.
22. Schwefel, H.-P. *Evolution and Optimum Seeking*; Wiley: New York, NY, USA, 1995.
23. Schumer, M.A.; Steiglitz, K. Adaptive Step Size Random Search. *IEEE Trans. Aut. Control* **1968**, *13*, 270–276. [CrossRef]
24. Rudolph, G. Local convergence rates of simple evolutionary algorithms with Cauchy mutations. *IEEE Trans. Evol. Comp.* **1997**, *1*, 249–258. [CrossRef]
25. Jägersküpper, J. Analysis of a simple evolutionary algorithm for minimisation in Euclidean spaces. In *International Colloquium on Automata, Languages, and Programming*; Lecture Notes in Computer Science; Springer: New York, NY, USA, 2003; Volume 2719, pp. 1068–1079.
26. Jägersküpper, J.; Witt, C. Rigorous runtime analysis of a $(\mu + 1)$ ES for the sphere function. In Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation, Washington, DC, USA, 25–29 June 2005; pp. 849–856.
27. Agapie, A.; Solomon, O.; Giuclea, M. Theory of (1+1) ES on the RIDGE. *IEEE Trans. Evol. Comp.* **2021**, *2021*, 3111232. [CrossRef]
28. Agapie, A.; Solomon, O.; Bǎdin, L. Theory of (1+1) ES on SPHERE revisited. **2021**, under review.
29. Beyer, H.-G. On the performance of $(1, \lambda)$-evolution strategies for the ridge function class. *IEEE Trans. Evol. Comput.* **2001**, *5*, 218–235. [CrossRef]
30. Cambanis, S.; Huang, S.; Simons, G. On the Theory of Elliptically Contoured Distributions. *J. Mult. Anal.* **1981**, *11*, 368–385. [CrossRef]
31. Huntington, E.V. Frequency Distribution of Product and Quotient. *Ann. Math. Statist.* **1939**, *10*, 195–198. [CrossRef]
32. Huang, H.; Su, J.; Zhang, Y.; Hao, Z. An Experimental Method to Estimate Running Time of Evolutionary Algorithms for Continuous Optimization. *IEEE Trans. Evol. Comput.* **2020**, *24*, 275–289. [CrossRef]
33. Abramowitz, M.; Stegun, I.A. (Eds.) *Handbook of Mathematical Functions*, 9th ed.; Dover: New York, NY, USA, 1972.
34. Neuman, E. Inequalities and Bounds for the Incomplete Gamma Function. *Results Math.* **2013**, *63*, 1209–1214. [CrossRef]
35. Volkmer, H.; Wood, J.J. A note on the asymptotic expansion of generalized hypergeometric functions. *Anal. Appl.* **2014**, *12*, 107–115. [CrossRef]
36. Slater, L.J. *Generalized Hypergeometric Functions*; University Press: Cambridge, UK, 1966.
37. Karp, D.B. Representations and Inequalities for Generalized Hypergeometric Functions. *J. Math. Sci.* **2015**, *207*, 885—897. [CrossRef]
38. Luke, Y.L. Inequalities for generalized hypergeometric functions. *J. Approx. Theory* **1972**, *5*, 41–65. [CrossRef]
39. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comp.* **1997**, *1*, 67–82. [CrossRef]
40. Kadirkamanathan, V.; Selvarajah, K.; Fleming, P.J. Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Trans. Evol. Comp.* **2006**, *10*, 245–255. [CrossRef]
41. Dasgupta, S.; Das, S.; Biswas, A.; Abraham, A. On stability and convergence of the population-dynamics in differential evolution. *AI Commun.* **2009**, *22*, 1–20. [CrossRef]