On Correlated Mutations in Evolution Strategies

Günter Rudolph[a]

[a]University of Dortmund, Department of Computer Science, P.O. Box 50 05 00,
D-W-4600 Dortmund 50

1. INTRODUCTION

Originally Evolution Strategies (ESs) have been developed for experimental optimization, i.e. optimization at the real object. Later they have been formulated as computer programs in order to solve optimization problems given as mathematical models of the type

$$\min\{f(\mathbf{x}) \,|\, \mathbf{x} \in M \subseteq \mathbb{R}^n\}\,, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is called the *objective function* and the set M the *feasible region*. If the objective function possesses several local minima the determination of the global or lowest local minimum is an unsolvable problem in general [13, p. 6]. In this paper we are interested mainly in convergence speed of ESs such that we shall assume for simplicity that f has only one local minimum which is of course the global one. First work on this topic has been done by [7], [9] and [8], who calculated the convergence rate of a stochastic process w.r.t. the problem

$$f_1(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x} = ||\mathbf{x}||^2 \quad , M = \mathbb{R}^n\,. \tag{2}$$

Rechenberg's so-called $(1+1)$-ES can be considered as one of the first approaches to use principles of biological evolution in the sense of optimization rules:

$$\mathbf{x}^{(t+1)} = \begin{cases} \mathbf{x}^{(t)} + \mathbf{z}^{(t)} & \text{, if } f(\mathbf{x}^{(t)} + \mathbf{z}^{(t)}) < f(\mathbf{x}^{(t)}) \text{ and } \mathbf{x}^{(t)} + \mathbf{z}^{(t)} \in M \\ \mathbf{x}^{(t)} & \text{, otherwise} \quad , \end{cases} \tag{3}$$

where $\mathbf{z}^{(t)} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ is a multinormally distributed random vector with zero mean and covariance matrix $\mathbf{C} = \sigma^2\mathbf{I}$. The addition of the random vector in (3) can be regarded as *mutation* and choosing the better point, if any, as *selection*. The ancestor here survives until an offspring surpasses him.

Various extensions have led to so-called multi-membered $(\mu + \lambda)$ and (μ, λ)-ESs with $1 \leq \mu < \lambda$ [11], where in the latter case μ parents produce λ offspring and the μ best of these will serve as new parents in the next generation. If $\mu > 1$, then the principle of recombination can be introduced. For a more detailed summary concerning the development of ESs see [1].

The most general form of a multinormal distribution with zero mean for the random vector $\mathbf{z}$ in (3) is $\mathbf{z} \sim N(\mathbf{0}, \mathbf{C})$, where $\mathbf{C}$ is any positive definite and symmetric correlation

matrix. The p.d.f. is given by

$$p_{\mathbf{z}}(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{x}\right)}{\sqrt{(2\pi)^n \det(\mathbf{C})}} \quad . \tag{4}$$

The advantage of using the general form instead of $\mathbf{C} = \sigma^2\mathbf{I}$ or $\mathbf{C} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ can be seen in the light of the second order Taylor series expansion of the objective function assuming — for the moment — that f is twice continuous differentiable:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^{\mathrm{T}}\nabla f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathrm{T}}\nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \quad . \tag{5}$$

Then, the structure of the Hessian matrix $\mathbf{H}(\mathbf{x}_0) := \nabla^2 f(\mathbf{x}_0)$ gives information about the topology of the objective function in a certain neighborhood of the point $\mathbf{x}_0$. For simplicity we give an example where the second order expansion is exact:

$$f_2(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} \quad \text{with } a_{ij} := n + 1 - \max\{i, j\} \text{ and } M = \mathbb{R}^n \quad . \tag{6}$$

Let $n = 2$. Then the Hessian for f_2 is $\mathbf{H}(\mathbf{x}) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$ for all $\mathbf{x} \in M$. The level sets $L_a := \{\mathbf{x} \in M : f(\mathbf{x}) = a\}$ of f_2 are ellipsoids whereas those of f_1 are spheres as can be seen from figure 1.

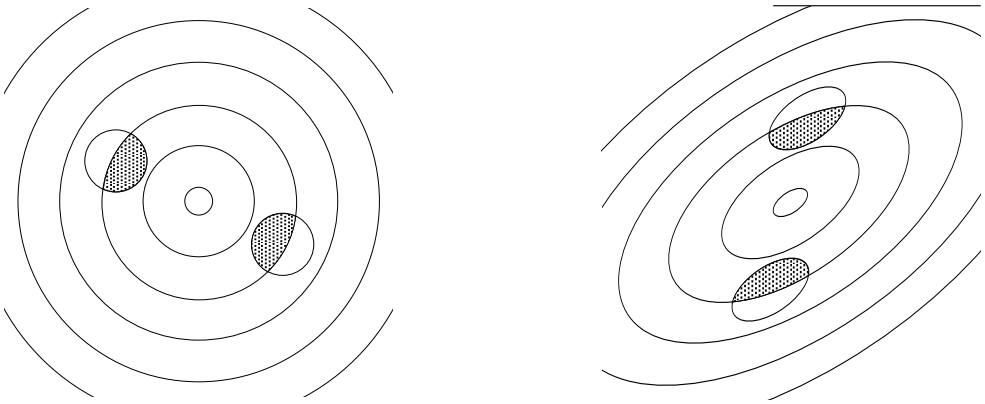

Figure 1: Level sets of f_1 (left) and f_2 (right) with regions of successful mutations (hatched areas)

The optimal distributions for both problems w.r.t. the covariance matrix are given if the alignment of the lines of equiprobable mutation steps is the same as the alignment of the level sets, because then the region of successful mutations (hatched areas in fig. 1) has the same size for any element of L_a (a fixed) chosen as center (parent). In other words, then

the elliptical problem f_2 is transformed to the simpler symmetric and spherical problem f_1. For a multinormal distribution this can be achieved by using the inverse of the Hessian as correlation matrix: $\mathbf{C} = \mathbf{H}^{-1}$. Now, multiplication of $\mathbf{C}$ with a suitable step size will speed up the convergence considerably compared to the more special correlation matrices mentioned above.

However, starting the ES without *a priori* knowledge w.r.t. the objective function the Hessian is unknown. In [12] Schwefel proposed the idea to approximate the inverse of the Hessian probabilistically simultaneous to the optimization process. Now the entries of the correlation matrix are mutated and recombined similar to the trial points $\mathbf{x}$. However, the problem is that we have to guarantee the positive definiteness and symmetry of $\mathbf{C}$ while the entries are modified probabilistically. Schwefel has given a geometric interpretation of his correlation routine but it is not obvious that this routine is capable of producing all feasible and only feasible correlation matrices.

In the next section we shall give the formal proof that the open question mentioned above can be answered positively. The proofs themselves will give some additional insight into the problem of constructing a positive definite and symmetric correlation matrix by means of a probabilistic adaptation process. Numerical results given in section 3 will show that an acceleration can be achieved but that it is well below the optimal limit with known Hessian. This leads to the question whether there is enough knowledge within the population in order to approximate $\mathbf{C}$ at all. In section 4 we shall give an upper bound on the population size such that the desired matrix can be obtained within one generation. Finally, we briefly discuss ESs with individual memory. This concept can be used to decrease the upper bound on the population size significantly.

2. CORRELATED MUTATIONS

In the following we shall derive the feasible ranges within $\frac{1}{2}n(n+1)$ coefficients may be chosen such that we can produce every positive definite and symmetric matrix. The first theorem gives the instruction how to produce a multinormal correlated random vector $\mathbf{x}$.

THEOREM 1:
Let $\mathbf{y} \sim N(\mathbf{0}, \mathbf{I})$ be a random vector of dimension n and $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} : n \times n$ a positive definite matrix. Then

$$\mathbf{x} \stackrel{d}{=} \mathbf{Q}^{\mathrm{T}}\mathbf{y} \sim N(\mathbf{0}, \mathbf{Q}^{\mathrm{T}}\mathbf{Q}). \tag{7}$$

Here, $\mathbf{x} \stackrel{d}{=} \mathbf{y}$ denotes that random vectors $\mathbf{x}$ and $\mathbf{y}$ have the same distribution.
PROOF: See e.g. [5, pp. 31–32]. □

In order to find a proper matrix $\mathbf{Q}$ we need the following

LEMMA 1:
Let $\mathbf{T} : n \times n$ be orthogonal (i.e. $\mathbf{T}^{\mathrm{T}} = \mathbf{T}^{-1}$) and $\mathbf{D} : n \times n$ a diagonal matrix with positive diagonal entries. Then

$$\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T} \tag{8}$$

is positive definite and symmetric.

PROOF:
$\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T} = \mathbf{T}^{-1}\mathbf{D}\mathbf{T}$. Thus, $\mathbf{A}$ and $\mathbf{D}$ are similar and have the same eigenvalues [3, p. 200]. Since all eigenvalues (the diagonal entries of $\mathbf{D}$) are positive, matrix $\mathbf{A}$ is positive definite.

$\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T} = \mathbf{T}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{T} = (\mathbf{T}\mathbf{S})^{\mathrm{T}}(\mathbf{T}\mathbf{S})$ with $\mathbf{S} = \mathbf{D}^{\frac{1}{2}}$. Now symmetry follows directly from matrix multiplication formula. □

Thus, $\mathbf{Q} = \mathbf{D}^{\frac{1}{2}}\mathbf{T}$. Next, we have to answer the question whether any positive matrix $\mathbf{A}$ can be decomposed by (8). This follows from the so–called principal axis theorem:

THEOREM 2:
Let $\mathbf{A}: n \times n$ be a symmetric and positive definite matrix. Then there exists an orthogonal matrix $\mathbf{T}: n \times n$ and a diagonal matrix $\mathbf{D}: n \times n$ with positive diagonal entries such that

$$\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T} \tag{9}$$

PROOF: See e.g. [3, pp. 271–275], [10, p. 112]. □

Then, $\mathbf{T}^{\mathrm{T}}$ is the so–called modal matrix of eigenvectors and the diagonal entries of $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$. This decomposition is unique except for permutations of the columns of $\mathbf{T}$ and the corresponding diagonal entries of $\mathbf{D}$. Thus, there are $n!$ matrices $\mathbf{T}$ and $\mathbf{D}$ which fulfill (9), if all eigenvalues are different. From the theory of determinants we know that for $\frac{1}{2}n!$ matrices $\det(\mathbf{T}) = +1$ and for the remaining matrices $\det(\mathbf{T}) = -1$.

Finally we have to give a decomposition of $\mathbf{T}$ such that every orthogonal matrix can be constructed which fulfills (9).

THEOREM 3:
Let $\mathbf{T}: n \times n$ be orthogonal with $\det(\mathbf{T}) = +1$. Then there exist $\frac{1}{2}n(n-1)$ angles $\omega_k \in (-\pi, \pi]$ such that $\mathbf{T}$ can be decomposed as a product of elementary rotation matrices $\mathbf{R}_{ij}(\omega_k)$:

$$\mathbf{T} = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \mathbf{R}_{ij}(\omega_k) \quad . \tag{10}$$

PROOF: See [4, p. 41]. □

An elementary rotation matrix $\mathbf{R}_{ij}(\omega)$ consists of the unit matrix except that for the following entries holds: $r_{ii} = r_{jj} = \cos(\omega), r_{ij} = -r_{ji} = -\sin(\omega)$. In [4] the transpose of $\mathbf{R}_{ij}(\omega_k)$ is used. However, exploiting the symmetry properties of the trigonometric functions we can give another set of angles which fulfill (10): $\mathbf{R}_{ij}(\omega_k) = \mathbf{R}^{\mathrm{T}}{}_{ij}(-\omega_k)$. The following lemma guarantees that there does not exist a set of angles ω_k such that (10) produces a matrix $\mathbf{T}$ that is not orthogonal.

LEMMA 2:
The product of elementary rotation matrices is orthogonal.

PROOF:
Every elementary rotation matrix is orthogonal: $\mathbf{R}^{\mathrm{T}}{}_{ij}\mathbf{R}_{ij} = \mathbf{I}$. Since a product of orthogonal matrices is orthogonal the proof is completed. □

Now we can summarize:

THEOREM 4:
A matrix $\mathbf{A}: n \times n$ is positive definite and symmetric iff $\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{T}$ with $\mathbf{T}$ as given in (10), $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_n)$ and $\omega_k \in (-\pi, \pi]$ for $k = 1, \ldots, \frac{1}{2}n(n-1)$ and $s_i > 0$ for $i = 1, \ldots, n$.

PROOF:
"$\Rightarrow$":
Let $\mathbf{A}: n \times n$ be positive definite and symmetric. By means of Theorem 2 there exist a decomposition $\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T}$, where $\mathbf{T}$ is orthogonal and $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_n)$ with $s_i > 0$. Suppose, this decomposition results in $\det(\mathbf{T}) = -1$. Since there are $\frac{1}{2}n!$ decompositions of a positive definite and symmetric matrix $\mathbf{A}: n \times n$ with $\det(\mathbf{T}) = +1$ and $\frac{1}{2}n!$ decompositions with $\det(\mathbf{T}) = -1$, we can simply exchange two columns in $\mathbf{T}$ and the corresponding diagonal entries of $\mathbf{D}$ such that for this decomposition holds $\det(\mathbf{T}) = +1$. That means, that there exists always a decomposition with $\det(\mathbf{T}) = +1$. Moreover, by Theorem 3 there exists a decomposition of $\mathbf{T}$ as given in (10) with angles $\omega_k \in (-\pi, \pi]$.
"$\Leftarrow$":
Let $\mathbf{R}_{ij}(\omega_k)$ be elementary rotation matrices with $\omega_k \in (-\pi, \pi]$. By Lemma 2 the product as given in (10) is orthogonal. Let $\mathbf{S}$ be a diagonal matrix with positive diagonal entries and $\mathbf{D} = \mathbf{S}^{\mathrm{T}}\mathbf{S}$. By Lemma 1 matrix $\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{D}\mathbf{T}$ is positive definite and symmetric. □

It is still left open which kind of probability distribution should be chosen for mutating the angles in the range $(-\pi, \pi]$ and the step sizes in the range $(0, \infty)$. The current implementation [12] employs an independent normal distribution for the angles

$$\omega_i^{(t+1)} = \left(\omega_i^{(t)} + z_\omega + \pi\right) \mod 2\pi - \pi \quad \text{with } z_\omega \sim N\left(0, \left(\frac{\kappa\pi}{180^\circ}\right)^2\right) \tag{11}$$

with $\kappa = 5^\circ$ and an independent lognormal distribution for the step size control parameters

$$\sigma_i^{(t+1)} = \sigma_i^{(t)} \cdot z_{\tilde{\sigma}} \cdot z_\sigma \quad \text{with } z_{\tilde{\sigma}} \sim LN(0, (2n)^{-1}),\ z_\sigma \sim LN(0, (4n)^{-\frac{1}{2}}) \ , \tag{12}$$

where $z_{\tilde{\sigma}}$ is realized only once for all i. We shall postpone this question and pick it up again when we discuss the numerical results of some test runs obtained with the above settings.

3. NUMERICAL RESULTS

In this section we intend to study the effectiveness of the probabilistic approximation procedure which is used to construct correlated random vectors. The test problem under consideration is f_2 as given in (6) with $n = 10$. Due to the parameter setting problem of Evolutionary Algorithms the simulation runs have been made with different recombination types using a $(\mu, 100)$-ES and varying μ from 2 to 30. The performance or progress measure of a variant has been defined as

$$P_{R,v} = \frac{1}{R}\sum_{i=1}^{R} \log_{10} \frac{f_{i,v}^{(0)}}{f_{i,v}^{(2000)}} \quad \text{with } R = 10 \ , \tag{13}$$

Table 1
Progress within 2000 generations sorted w.r.t. the recombination type of object variables x, step sizes s and angles a with optimal μ

object variables			
v	1	2	3
x11	84	109	108
x12	106	122	117
x13	70	97	71
x21	100	133	130
x22	123	135	138
x23	77	108	85
x31	140	150	160
x32	150	153	166
x33	133	141	140

step sizes			
v	1	2	3
1s1	84	100	140
1s2	106	123	150
1s3	70	77	133
2s1	109	133	150
2s2	122	135	153
2s3	97	108	141
3s1	108	130	160
3s2	117	138	166
3s3	71	85	140

angles			
v	1	2	3
11a	84	106	70
12a	100	123	77
13a	140	150	133
21a	109	122	97
22a	133	135	108
23a	150	153	141
31a	108	117	71
32a	130	138	85
33a	160	166	140

where R denotes the number of runs and $f_{i,v}^{(t)}$ the objective value of version v at generation t in the i-th run. The version $v \in \{1,2,3\}^3$ can be identified by three digits with the following meaning: the position of the digits is associated with the recombination type of the object variables x, step size control parameters s and angles a: $v = xsa$. The digit value 1 expresses that no recombination is used whereas digit values 2 and 3 represent discrete and intermediate recombination, respectively. Using discrete recombination each entry of the offspring vector (gene) is chosen either from the first or second parent's vector with the same probability, whereas by using intermediate recombination the parents' vectors are averaged, e.g. $x_{child} = \frac{1}{2}(x_{parent1} + x_{parent2})$.

The resulting progress rates are summarized in table 1, where the progress of each variant is recorded for the best choice of the number of parents μ. Independently of the recombination type the best results are obtained if μ is chosen between 4 and 6. A closer look at table 1 reveals some interesting facts: The second column of the right table is dominating the other two indicating that discrete recombination of the angles is preferable for this test problem. Another dominance can be noticed for the step sizes: Here, intermediate recombination dominates the other variants. Consequently, the row with the entry $x32$ dominates all other rows in the left table. The progress together with the standard deviation over 10 runs for the best variant $v = 332$ is given in figure 2. Although the peak performance of about 166 orders of magnitude appears to be promising the performance with complete knowledge (with known Hessian, resp.) is above 500 orders of magnitude, as we shall see later.

Indeed, the correlation procedure seems to work but it achieves only a third of what is possible for our test problem. There are some imaginable reasons for the partial failure.

First, recall the correlation procedure:

$$\mathbf{Q} = \mathbf{S} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \mathbf{R}_{ij}(\omega_k) \quad . \tag{14}$$

Let be $n = 3$. Then there are three different sets of angles which build the same positive definite matrix $\mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ if the diagonal entries of $\mathbf{S}$ are permuted in a suitable

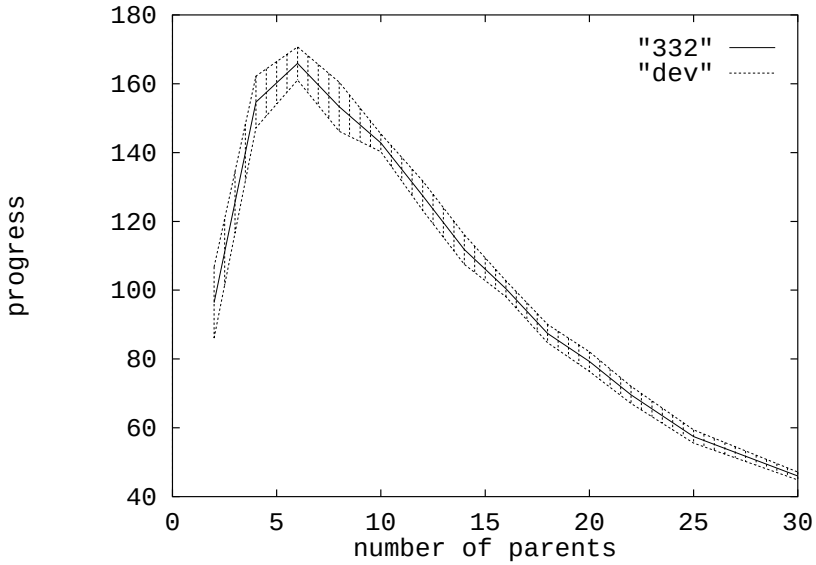

Figure 2: Progress with error bounds of the best variant

way. On the other hand, if the diagonal entries of **S** (the step sizes) will be permuted although the angles are well adapted already, then the angles have to be adapted anew. Such a permutation can occur easily due to the multiplication with the lognormal random variable, see (12). For example, if **A** as given in (6) is the desired matrix, then there are the following triples of angles w.r.t. the permutation of the step sizes: $(0.676, 0.334, -0.676), (1.990, -0.632, 1.152)$, and $(-0.507, 2.313, -0.507)$. That is why the matrix building process seems to be unstable in the sense that the adaptation of angles is perturbed by the adaptation process of the step sizes.

Second, the choice of a normal distribution with fixed variance as given in (11) might have another negative effect on the adaptation process. Again, suppose the angles are well adapted for a moment such that only the n step sizes are to be adjusted. However, due to the fixed variance mentioned above the angles will be modified again and again preventing the angle adaptation process to become stable.

Third, it is not obvious why the distributions used should be the right choice. However, before one can answer this question theoretically a lot of basic research has to be done. Another possibility for why the approximation procedure is not as effective as desired might consist in a lack of information within the population. In the next section we shall give an upper bound w.r.t. the population size such that the Hessian can be calculated deterministically if the knowledge of the whole population is exploited.

4. UPPER BOUNDS ON THE POPULATION SIZE

We shall state the main result of this section in the following theorem and give a constructive proof. Although this method can be used to determine an estimator for the

Hessian matrix, it is not recommended to do this due to the computational effort required.

THEOREM 5:
If the objective function has a positive definite Hessian matrix $\mathbf{H} : n \times n$ in a certain neighborhood of a point $\mathbf{x}_0 \in M$, then a unique least squares estimator of the Hessian matrix can be constructed with not more than $\frac{1}{2}(n^2 + 3n + 4)$ pairs $(\mathbf{x}, f(\mathbf{x}))$.

PROOF:
The second order Taylor expansion (5) of f can be formulated as

$$T_2(\mathbf{x}; \mathbf{x}_0) = b_0 + \mathbf{b}^{\mathrm{T}}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{B}\mathbf{x} \quad . \tag{15}$$

The unknown entries $(b_0, \mathbf{b}, \mathbf{B})$ can be gathered in a vector

$$\mathbf{c}^{\mathrm{T}} = (c_1, \ldots, c_N) := (b_0, b_1, \ldots, b_n, b_{11}, b_{12}, \ldots, b_{1n}, b_{22}, b_{23}, \ldots, b_{n-1,n-1}, b_{n-1,n}, b_{nn})$$

with $N = \frac{1}{2}(n^2 + 3n + 2)$ due to the symmetry of $\mathbf{B}$ in (15). The least square estimator $\hat{\mathbf{c}}$ of $\mathbf{c}$ is unique, if we have $N + k$ pairs of samples $(\mathbf{x}, f(\mathbf{x}))$ which form a system of linear equations $\mathbf{G}\mathbf{c} = \mathbf{f}$ with $\mathbf{G}^{\mathrm{T}} = (\mathbf{g}_1, \ldots, \mathbf{g}_{N+k}), k > 0$, where

$$\begin{aligned} \mathbf{g}^{\mathrm{T}} &= (g_1, \ldots, g_N) \\ &:= (1, x_1, \ldots, x_n, x_1^2, 2x_1x_2, \ldots, 2x_1x_n, x_2^2, 2x_2x_3, \ldots, 2x_2x_n, \ldots, x_{n-1}^2, 2x_{n-1}x_n, x_n^2) \end{aligned}$$

and $\mathbf{f}^{\mathrm{T}} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_{N+k}))$. Moreover, matrix $\mathbf{G}^{\mathrm{T}}\mathbf{G}$ is required to be regular (see e.g. [10, p. 82-84]). Now, $\hat{\mathbf{c}} = (\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{f}$. From $\hat{\mathbf{c}}$ we select the entries $\hat{c}_{n+2}, \ldots, \hat{c}_N$ as entries for our estimator $\hat{\mathbf{B}}$ of the Hessian. Since more than N pairs $(\mathbf{x}, f(\mathbf{x}))$ are necessary, $N + 1 = \frac{1}{2}(n^2 + 3n + 4)$ is an upper bound. □

For numerical reasons $\hat{\mathbf{B}}$ should be normalized, such that $\hat{\mathbf{H}} := (\det(\hat{\mathbf{B}}))^{1/n}\ \hat{\mathbf{B}}$ with $\det(\hat{\mathbf{H}}) = 1$. Now, by Cholesky decomposition of $\hat{\mathbf{H}}$ and inverting the resulting triangular matrix we get the desired matrix $\mathbf{Q}$ to construct correlated random vectors via (7). The above procedure has complexity $O(N^3) = O(n^6)$ which is prohibitive for practical use. Nevertheless, the results of the test implementation demonstrate the effectiveness of the method (see fig. 3).

It follows from theorem 5 that $\mu+\lambda = \frac{1}{2}(n^2+3n+4)$ is an upper bound on the population size such that an approximation of the Hessian can be computed at each generation with the knowledge of the current population. This bound can be decreased from $O(n^2)$ to $O(n)$ if we introduce individuals with memory. For example, let the individuals store their phylogenetic tree up to a depth of n, then the bound decreases to $O(n)$. This is a hint to think again about the current implementations of Evolutionary Algorithms: One might make better use of the information gathered by the parents instead of throwing it away after their offspring are created.

5. CONCLUSIONS

In this paper it has been shown that the probabilistic approximation procedure of [12] can be used to construct any valid correlated multinormal random vector. Numerical results, however, indicate that the convergence of the approximation is not yet satisfactory.

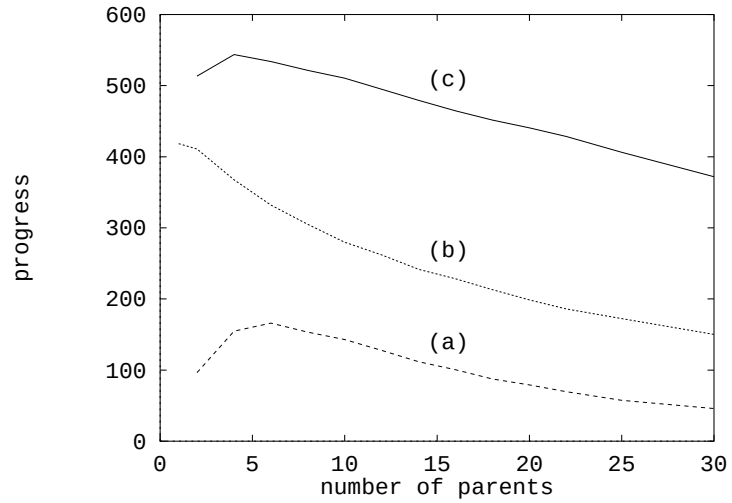

Figure 3: Progress of best variant with adapted correlation matrix (a), deterministic variant without recombination (b) and with recombination (c)

The main reason might be found in the fact that the step size adaptation process affects the angle adaptation process in a disruptive way. Thus, the approximation procedure has to be modified.

The above problem can be avoided by the proposed deterministic method which is used to construct the desired correlation matrix. Although numerical results demonstrate the effectiveness of this method it is of no practical use due to its time complexity. It has not been the intention to recommend this method but to gain an upper bound on the population size such that the desired correlation matrix can be obtained within one generation. There are other methods which can be used for this problem, for example any approximation routine of the variable metric method (see [6]). Such ideas immediately lead to the concept of individuals with memory which can make ESs even more powerful.

One might argue that a variable metric type method should be used whenever it is possible to approximate the Hessian. Surely, on the one hand ESs will converge more slowly but on the other hand they provide the probabilistic guarantee to find the global minimum of a multimodal problem contrary to a variable metric method. Such theoretical properties paired with the inherent parallelism make ESs so attractive for practical use.

6. ACKNOWLEDGEMENT

The author is supported by the Research Initiative *Parallel Computing* of Nordrhein–Westfalen, Land of the Federal Republic of Germany (project 107 004 91).

7. REFERENCES

1 Th. Bäck, F. Hoffmeister, and H.-P. Schwefel. A survey of evolution strategies. In

Belew and Booker [2], pages 2–9.

2 R.K. Belew and L.B. Booker, editors. *Proceedings of the fourth Conference of Genetic Algorithms*. Morgan Kaufmann, San Mateo, 1991.

3 C.W. Curtis. *Linear Algebra*. Undergraduate texts in mathematics. Springer, New York, 2nd edition, 1984.

4 D.K. Faddejew and W.N. Faddejewa. *Numerische Methoden der Linearen Algebra*. Oldenbourg, München, 3rd edition, 1973.

5 K.-T. Fang, S. Kotz, and K.-W. Ng. *Symmetric Multivariate and Related Distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London and New York, 1990.

6 W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes*. Cambridge University Press, Cambridge, 1986.

7 L.A. Rastrigin. The convergence of the random search method in the extremal control of a many-parameter system. *Automation and Remote Corntrol*, 24:1337–1342, 1963.

8 Ingo Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann–Holzboog Verlag, Stuttgart, 1973.

9 M.A. Schumer and K. Steiglitz. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 13:270–276, 1968.

10 H.R. Schwarz, H. Rutishauser, and E. Stiefel. *Numerik symmetrischer Matrizen*, volume 11 of *Leitfäden der angewandten Mathematik und Mechanik*. Teubner, Stuttgart, 1968.

11 Hans-Paul Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Interdisciplinary systems research; 26. Birkhäuser, Basel, 1977.

12 Hans-Paul Schwefel. *Numerical Optimization of Computer Models*. Wiley, Chichester, 1981.

13 A. Törn and A. Zilinskas. *Global Optimization*, volume 350 of *Lecture Notes in Computer Science*. Springer, Berlin and Heidelberg, 1989.