Optimal Weighted Recombination

Dirk V. Arnold

Technical Report CS-2004-08

August 9, 2004

Faculty of Computer Science
6050 University Ave., Halifax, Nova Scotia, B3H 1W5, Canada

Abstract. Weighted recombination is a means for improving the local search performance of evolution strategies. It aims to make effective use of the information available, without significantly increasing computational costs per time step. In this paper, the potential speed-up resulting from using rank-based weighted recombination is investigated. Optimal weights are computed for the sphere model, and comparisons with the performance of strategies that do not make use of weighted recombination are presented. It is seen that unlike strategies that rely on unweighted recombination and truncation selection, weighted multirecombination evolution strategies are able to improve on the serial efficiency of the $(1+1)$-ES on the sphere. The implications of the use of weighted recombination for noisy optimization are studied, and parallels to the use of rescaled mutations are drawn. The cumulative step length adaptation mechanism is formulated for the case of an optimally weighted evolution strategy, and its performance is analyzed.

1 Introduction

In his seminal book Rechenberg [17] in 1973 presented the derivation of a law describing the progress rate of the $(1+1)$-ES on the high-dimensional sphere model. From that law, it can be seen numerically that for optimally adapted mutation strength, the normalized rate at which the optimum is approached equals 0.202. In the years that followed, evolution strategies evolved. The single-parent strategy was replaced by population-based strategies, and recombination was introduced. In 1996, Beyer [8] studied the performance on the sphere model of the $(\mu/\mu, \lambda)$-ES — a population-based strategy that uses multi-recombination. He made the surprising discovery that the serial efficiency of the $(\mu/\mu, \lambda)$-ES for optimally chosen population size parameters asymptotically approaches the same value of 0.202 that the $(1+1)$-ES had achieved more than two decades earlier. Moreover, while few theoretical results exist, there is evidence that none of the $(\mu/\rho \overset{+}{,} \lambda)$-ES achieve a serial efficiency on the sphere model that exceeds that of the simple $(1+1)$-ES. Needless to say, this is not to imply that no progress had been made. Population-based strategies allow for parallelization, have greater adaptation capabilities, and are much superior when applied to noisy optimization problems. Nonetheless, the $(1+1)$-ES sets the benchmark for serial efficiency on the simple sphere model.

Key to achieving a serial efficiency that exceeds that of the $(1+1)$-ES is to recognize that generally, all $(\mu/\rho \overset{+}{,} \lambda)$-ES discard information. Truncation selection leads to all of the selected offspring having the same influence on the progress of the strategy, irrespective of their respective ranks within the population. For example, for the $(\mu/\mu, \lambda)$-ES the influence of the best candidate solution equals that of the μth best. Similarly, all relative rank information from those offspring that are not selected to survive is discarded. Those candidate solutions are without influence on the step taken by the strategy, no matter whether they narrowly missed the cut or they missed it by a wide margin.

More complete use of the information gained by evaluating offspring candidate solutions can be made by weighting their influence in the recombination and selection process. Weights can be chosen such that they more carefully discriminate between good and bad candidate solutions than truncation selection does. The choice of weights can be based either on function values or on rank within the set of offspring generated. A strategy that uses function values to determine weights is the evolutionary gradient search strategy (EGS) proposed by Salomon [20]. EGS differs from evolution strategies not only in its reliance on function values rather than ordinal data, but also in its use of "negative information". The weight assigned to a candidate solution is proportional to the difference between that candidate solution's fitness and the fitness of the search point that it has been generated from. As a consequence, those offspring that improve on the previous time step's fitness receive positive weights, and offspring that are inferior to their parent receive negative weights and thus result in the strategy moving in the opposite direction. An investigation in [2] has shown that EGS is indeed capable of achieving serial efficiencies on the sphere model that exceed those of the $(1+1)$-ES. However, it has also been seen that the explicit rescaling of progress vectors that EGS performs hampers genetic repair, and that as a result EGS is generally inferior to the $(\mu/\mu, \lambda)$-ES in the presence of noise as well as if implemented on parallel computers.

Rank-based weighted recombination has been employed by Hansen and Ostermeier [14] in connection with their covariance matrix adaptation evolution strategy (CMA-ES), and it has also been used in the comparative review of evolutionary algorithms by Kern et al. [15]. In both references, it is suggested to assign positive weights of different magnitudes to the better 50 percent of the candidate solutions generated. A heuristic rule for choosing those weights is proposed. Without a reason being given, but probably in realization of the fact that the opposite of a bad direction is not always a good direction, the use of negative weights is discouraged. Zero weights are assigned to the inferior 50 percent of candidate solutions generated. In [14] it is noted that speed-up factors of less than two are observed compared to the $(\mu/\mu, \lambda)$-ES. A direct and systematic comparison between weighted and unweighted recombination is not performed.

The only attempt made so far to explore the consequences of the choice of weights analytically has been made by Rudolph [19]. For a weighted strategy that generates offspring by placing them on a sphere shell rather than by Gaussian mutations, Rudolph computes expressions for the progress rate on the sphere model. Those expressions involve expectations of joint beta order statistics and are difficult to determine in the general case. For that reason, Rudolph explores consequences of his results only for the case that the search space dimensionality equals three. Even for this special case the resulting expressions are too complicated to determine optimal weights. Rudolph does observe that the use of negative weights can have effects beneficial for the progress rate of the strategy, and that the serial efficiency of a strategy using unweighted recombination in connection with truncation selection can be exceeded by strategies that make use of weighted recombination.

It is the goal of this paper to obtain an improved understanding of the interplay of mutation, recombination, and selection in evolution strategies, and of the potential that weighted recombination has to speed up local search. In contrast to the aforementioned paper by Rudolph, focus here is on Gaussian mutations and the case that the search space dimensionality is high. The situation has the advantage of being comparatively well understood, and of allowing an analytical treatment. The results obtained are exact only in the limit of infinite search space dimensionality, but they do contribute to the understanding of the evolutionary processes in sufficiently high-dimensional spaces.

The remainder of this paper is organized as follows. The sphere model as an important environment for studying local search properties of direct optimization strategies as well as weighted multirecombination evolution strategies are introduced in Sect. 2. In Sect. 3, the quality gain of weighted multirecombination evolution strategies is computed and optimal weights are determined. Section 4 addresses the issue of how the performance of evolution strategies with optimally weighted multirecombination is affected by noise. It is seen that the issue of rescaled mutations raised by Rechenberg [18] and studied by Beyer [9, 10] arises naturally in connection with the choice of weights and the issue of genetic repair in multirecombination strategies. In Sect. 5, the cumulative step length adaptation mechanism is formulated for the case of the optimally weighted multirecombination evolution strategy, and its performance is analyzed. Finally, Sect. 6 concludes with a brief summary and directions for future research.

2 Preliminaries

In this section evolution strategies using weighted multirecombination for the minimization of functions $f : \mathbb{R}^N \to \mathbb{R}$ are formally introduced. Then the sphere model is briefly discussed as an important environment for learning about the behavior of local search algorithms.

2.1 Weighted Multirecombination Evolution Strategies

Weighted multirecombination evolution strategies repeatedly update a search point $\mathbf{x} \in \mathbb{R}^N$ using the following four steps:

1. Generate λ offspring candidate solutions $\mathbf{y}^{(i)} = \mathbf{x} + \sigma\mathbf{z}^{(i)}$, $i = 1, \ldots, \lambda$. The $\mathbf{z}^{(i)}$ are vectors consisting of N independent, standard normally distributed components and are referred to as mutation vectors. The nonnegative quantity σ is referred to as the mutation strength and determines the step length of the strategy.
2. Determine the objective function values $f(\mathbf{y}^{(i)})$ of the offspring candidate solutions and order the $\mathbf{y}^{(i)}$ according to those values. After ordering, index $k;\lambda$ refers to the kth best of the λ offspring (the kth smallest for minimization; the kth largest for maximization).
3. Compute the weighted average
$$\mathbf{z}^{(\mathrm{avg})} = \sum_{k=1}^{\lambda} w_{k;\lambda}\mathbf{z}^{(k;\lambda)} \tag{1}$$
of the $\mathbf{z}^{(i)}$ vectors. The $w_{k;\lambda}$ are weights that depend on the rank of the corresponding candidate solution in the set of all offspring.
4. Replace the search point $\mathbf{x}$ by $\mathbf{x} + \sigma\mathbf{z}^{(\mathrm{avg})}$.

The vector $\mathbf{z}^{(\mathrm{avg})}$ defined in Eq. (1) is referred to as the progress vector. Clearly, $\sigma\mathbf{z}^{(\mathrm{avg})}$ connects consecutive search points. Notice that for the particular choice of weights
$$w_{k;\lambda} = \begin{cases} 1/\mu & \text{if } 1 \leq k \leq \mu \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$
the weighted multirecombination evolution strategy simply is the $(\mu/\mu, \lambda)$-ES. In that case, the search point $\mathbf{x}$ is the centroid of the population that consists of the μ best of the λ offspring candidate solutions generated. Also notice that the evolutionary gradient search strategy introduced in [20] does not entirely fit into

the framework of rank-based weighted multirecombination as weights are chosen proportional to $f(\mathbf{x}) - f(\mathbf{y}^{(i)})$ rather than based on rank, and as a normalization step is required between the averaging of mutation vectors and the update of the search point.

2.2 The Sphere Model

Since the early work of Rechenberg [17], the local performance of evolution strategies has commonly been studied on the quadratic sphere given by objective function

$$f(\mathbf{x}) = (\hat{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}(\hat{\mathbf{x}} - \mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^N,$$

where the task is minimization and where $\hat{\mathbf{x}} \in \mathbb{R}^N$ is the optimizer. The sphere serves as a model for objective functions in the vicinity of well-behaved local optima. See [5] for a justification of the usefulness of such considerations and for possible generalizations. Perhaps most importantly, strategies such as the CMA-ES described in [14] have been found to effectively scale a wide range of convex quadratic functions into the sphere, opening up the possibility that findings made for the sphere model have much wider-ranging significance.

In order to quantify the local performance of search strategies on the sphere, consider the effect of adding a vector $\sigma\mathbf{z}$ to the current search point $\mathbf{x}$. Multirecombination evolution strategies do so both when generating offspring candidate solutions and when updating the search point at the end of an iteration. Denoting the respective distances of $\mathbf{x}$ and $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$ from the optimizer by R and r, the difference $\delta(\mathbf{z}) = R^2 - r^2$ between objective function values $f(\mathbf{x}) = R^2$ and $f(\mathbf{y}) = r^2$ is referred to as the fitness advantage associated with vector $\mathbf{z}$.[1] The fitness advantage associated with mutation vectors determines the ordering of the candidate solutions and thus the weights with which those mutation vectors enter recombination. The fitness advantage associated with the progress vector can be used for defining a performance measure for evolution strategies as seen below.

The commonly used approach to determining $\delta(\mathbf{z})$ on the sphere model relies on a decomposition of vector $\mathbf{z}$ that has been used in [11, 18] and that is illustrated in Fig. 1. A vector $\mathbf{z}$ originating at search space location $\mathbf{x}$ can be written as the sum of two vectors $\mathbf{z}_A$ and $\mathbf{z}_B$, where $\mathbf{z}_A$ is parallel to $\hat{\mathbf{x}} - \mathbf{x}$ and $\mathbf{z}_B$ is in the $(N-1)$-dimensional hyperplane perpendicular to that. The vectors $\mathbf{z}_A$ and $\mathbf{z}_B$ are referred to as the central and lateral components of vector $\mathbf{z}$, respectively. The signed length z_A of the central component of vector $\mathbf{z}$ is defined to equal $\|\mathbf{z}_A\|$ if $\mathbf{z}_A$ points towards the optimizer and to equal $-\|\mathbf{z}_A\|$ if it points away from it. Using elementary geometry, it can easily be seen that

$$r^2 = (R - \sigma z_A)^2 + \sigma^2\|\mathbf{z}_B\|^2,$$

[1] While the notation adopted here is deliberately brief and does not reflect that explicitly, it is important to keep in mind that the fitness advantage $\delta(\mathbf{z})$ depends not only on vector $\mathbf{z}$ but also on the mutation strength σ.

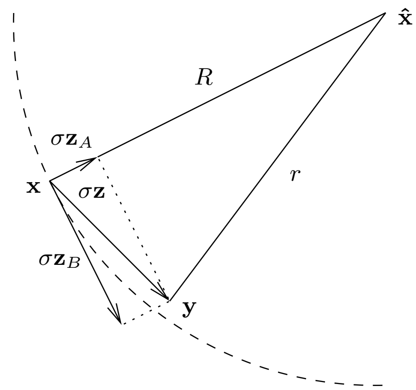

Fig. 1. *Decomposition of a vector $\mathbf{z}$ into central component $\mathbf{z}_A$ and lateral component $\mathbf{z}_B$. Vector $\mathbf{z}_A$ is parallel to $\hat{\mathbf{x}} - \mathbf{x}$, vector $\mathbf{z}_B$ is in the hyperplane perpendicular to that. The starting and end points, $\mathbf{x}$ and $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$, of vector $\sigma\mathbf{z}$ are at distances R and r from the optimizer $\hat{\mathbf{x}}$, respectively.*

and therefore, rearranging terms and noticing that $\|\mathbf{z}\|^2 = z_A^2 + \|\mathbf{z}_B\|^2$, that

$$\begin{aligned}\delta(\mathbf{z}) &= R^2 - r^2 \\ &= 2R\sigma z_A - \sigma^2\|\mathbf{z}\|^2.\end{aligned}$$

Introducing normalized quantities

$$\sigma^* = \sigma\frac{N}{R} \quad \text{and} \quad \delta^* = \delta\frac{N}{2R^2},$$

it follows

$$\delta^*(\mathbf{z}) = \sigma^* z_A - \frac{{\sigma^*}^2}{2N}\|\mathbf{z}\|^2 \tag{3}$$

for the normalized fitness advantage associated with vector $\mathbf{z}$.

In order to compute the normalized fitness advantage associated with vector $\mathbf{z}$ using Eq. (3), both the squared length and the signed length of the central component of that vector need to be determined. For the case that $\mathbf{z}$ is a mutation vector, it is well known from [11] that z_A is standard normally distributed, and that $\lim_{N\to\infty} \mathrm{E}[\|\mathbf{z}\|^2]/N = 1$. Furthermore, the variance of $\|\mathbf{z}\|^2/N$ tends to zero as N increases, making it possible to approximate $\|\mathbf{z}\|^2/N$ with unity provided that N is sufficiently large. Thus, the normalized fitness advantage associated with a mutation vector is asymptotically normally distributed with mean $-{\sigma^*}^2/2$ and with variance ${\sigma^*}^2$.

A commonly used performance measure for local search strategies is the quality gain which measures the rate at which the optimum is approached in the space of fitness values. It is defined as the expectation of the normalized fitness advantage associated with the progress vector and is thus

$$\begin{aligned}\varDelta^* &= \mathrm{E}\left[\delta^*\left(\mathbf{z}^{(\mathrm{avg})}\right)\right] \\ &= \sigma^*\mathrm{E}\left[z_A^{(\mathrm{avg})}\right] - \frac{{\sigma^*}^2}{2N}\mathrm{E}\left[\|\mathbf{z}^{(\mathrm{avg})}\|^2\right]. \end{aligned} \tag{4}$$

Another common performance measure — the progress rate — measures the rate at which the optimizer is approached in search space and is known from [11] to agree asymptotically with the quality gain on the sphere model for high search space dimensionality provided that appropriate normalizations are used. Moreover, as a performance measure that takes computational costs into account, it is commonplace to define the serial efficiency η of evolution strategies as the maximal quality gain per evaluation of the objective function. As the number of objective function evaluations per time step is λ, the serial efficiency of an evolution strategy is

$$\eta = \frac{1}{\lambda}\max_{\sigma^*}\varDelta^*. \tag{5}$$

Inherent in this definition are the assumptions that computational costs are dominated by the cost of evaluating the fitness function, and that evaluations need to be performed one after the other on a single processor.

3 Optimal Weighted Recombination

In this section an expression for the quality gain of the weighted multirecombination evolution strategy on the sphere model is derived that generalizes the corresponding result for the $(\mu/\mu, \lambda)$-ES obtained in [11, 18]. Then, optimal settings for the mutation strength and the recombination weights are computed, and consequences for the quality gain of the strategy are discussed.

3.1 Determining the Quality Gain

In order to determine the quality gain of the weighted multirecombination evolution strategy using Eq. (4), expected values of the squared length and of the signed length of the central component of the progress vector defined in Eq. (1) need to be computed. The progress vector's squared length is

$$\begin{aligned}\|\mathbf{z}^{(\mathrm{avg})}\|^2 &= \sum_{i=1}^{N}\left(\sum_{k=1}^{\lambda} w_{k;\lambda} z_i^{(k;\lambda)}\right)^2 \\ &= \sum_{i=1}^{N}\sum_{k=1}^{\lambda} w_{k;\lambda}^2 {z_i^{(k;\lambda)}}^2 + \sum_{i=1}^{N}\sum_{k\neq l} w_{k;\lambda} w_{l;\lambda} z_i^{(k;\lambda)} z_i^{(l;\lambda)}, \end{aligned} \tag{6}$$

where the $z_i^{(j)}$ are the components of the mutation vectors and as such standard normally distributed. The second term on the right hand side is a crosstalk term with mean zero. Thus, taking the expectation and exchanging the order of the summations in the first term it follows that

$$\begin{aligned}\frac{\mathrm{E}\left[\|\mathbf{z}^{(\mathrm{avg})}\|^2\right]}{N} &= \sum_{k=1}^{\lambda} w_{k;\lambda}^2 \frac{\mathrm{E}\left[\|\mathbf{z}^{(k;\lambda)}\|^2\right]}{N} \\ &\overset{N\to\infty}{=} \sum_{k=1}^{\lambda} w_{k;\lambda}^2, \end{aligned} \tag{7}$$

where in the second step we have made use of the important fact noted in Sect. 2 that asymptotically, $\mathrm{E}[\|\mathbf{z}\|^2]/N \to 1$ for mutation vector $\mathbf{z}$.

As for the expected signed length of the central component of the progress vector, it follows from the definition of that vector in Eq. (1) that

$$\mathrm{E}\left[z_A^{(\mathrm{avg})}\right] = \sum_{k=1}^{\lambda} w_{k;\lambda} \mathrm{E}\left[z_A^{(k;\lambda)}\right],$$

where of course $z_A^{(k;\lambda)}$ is the signed length of the central component of the mutation vector that corresponds to the kth best offspring candidate solution. In order to compute the expectations, it is important to recall from Sect. 2 that the $z_A^{(i)}$ are standard normally distributed. Moreover, neglecting the fluctuations of the crosstalk term in Eq. (6) (that can be seen to decrease in importance as N increases), the signed lengths of the central components of the mutation vectors determine the fitness of the corresponding offspring candidate solutions in that the offspring with the kth largest value of z_A is the kth fittest. Thus, in the limit of infinite search space dimensionality, $z_A^{(k;\lambda)}$ is the $(\lambda+1-k)$th order statistic of a sample of λ independent realizations of a standard normally distributed random variate. According to [7], the probability density function of $z_A^{(k;\lambda)}$ is

$$p_{k;\lambda}(x) = \frac{1}{\sqrt{2\pi}} \frac{\lambda!}{(\lambda-k)!(k-1)!} \mathrm{e}^{-\frac{1}{2}x^2} [\Phi(x)]^{\lambda-k} [1-\Phi(x)]^{k-1}, \tag{8}$$

where $\Phi(x)$ denotes the cumulative distribution function of the standardized normal distribution. It thus follows that the expected value of the signed length of the central component of the progress vector is

$$\mathrm{E}\left[z_A^{(\mathrm{avg})}\right] \overset{N\to\infty}{=} \sum_{k=1}^{\lambda} w_{k;\lambda} E_{k;\lambda}, \tag{9}$$

where

$$E_{k;\lambda} = \mathrm{E}\left[z_A^{(k;\lambda)}\right] = \int_{-\infty}^{\infty} x p_{k;\lambda}(x) \mathrm{d}x$$

denotes the expectation of the $(\lambda+1-k)$th order statistic and can easily be obtained by numerical integration.

Using Eqs. (7) and (9) in Eq. (4), it follows that the quality gain of the weighted multirecombination evolution strategy is

$$\Delta^* \stackrel{N\to\infty}{=} \sigma^* \sum_{k=1}^{\lambda} w_{k;\lambda} E_{k;\lambda} - \frac{{\sigma^*}^2}{2} \sum_{k=1}^{\lambda} w_{k;\lambda}^2. \tag{10}$$

Note that for the choice of weights in Eq. (2), Eq. (10) agrees with the quality gain law for the $(\mu/\mu, \lambda)$-ES derived in [11, 18].

3.2 Optimal Parameter Settings

Of course, it is desirable to choose the strategy's parameters such that the quality gain is maximized. Demanding that the derivative of Eq. (10) with respect to σ^* equals zero yields optimal normalized mutation strength

$$\sigma^* = \frac{\sum_{k=1}^{\lambda} w_{k;\lambda} E_{k;\lambda}}{\sum_{k=1}^{\lambda} w_{k;\lambda}^2}. \tag{11}$$

Reinserting this result into Eq. (10), the quality gain of the strategy for optimally adapted mutation strength is

$$\Delta^* = \frac{1}{2} \frac{\left(\sum_{k=1}^{\lambda} w_{k;\lambda} E_{k;\lambda}\right)^2}{\sum_{k=1}^{\lambda} w_{k;\lambda}^2}. \tag{12}$$

Optimal weights $w_{k;\lambda}$ can now be determined by computing the derivatives of Eq. (12) with respect to $w_{k;\lambda}$ for $k = 1, \ldots, \lambda$. Demanding that all derivatives be zero yields the system of equations

$$E_{k;\lambda} \sum_{l=1}^{\lambda} w_{l;\lambda}^2 = w_{k;\lambda} \sum_{l=1}^{\lambda} w_{l;\lambda} E_{l;\lambda}, \qquad k = 1, \ldots, \lambda. \tag{13}$$

Clearly, the system can be solved by setting $w_{k;\lambda} = E_{k;\lambda}$ for $k = 1, \ldots, \lambda$, and it is easily seen that the corresponding extremum really is a maximum. Therefore, optimal weights of the multirecombination evolution strategy on the sphere model are given by the first moments of the order statistics of the standardized normal distribution. We will refer to the strategy with optimally chosen weights as $(\lambda)_{\text{opt}}$-ES.

The dependence of the optimal weights on the rank within the set of candidate solutions is illustrated in Fig. 2. It can be seen that in order to achieve maximal progress on the sphere model, half of the offspring should enter recombination with positive weights, the other half should receive negative weights. Optimal weights are symmetric in that for every positive weight, there is a negative weight of equal value. This is in contrast to the behavior of EGS that assigns negative weights to the majority of the offspring generated as in convex environments, most will be inferior to their parent. Also note that the curves in

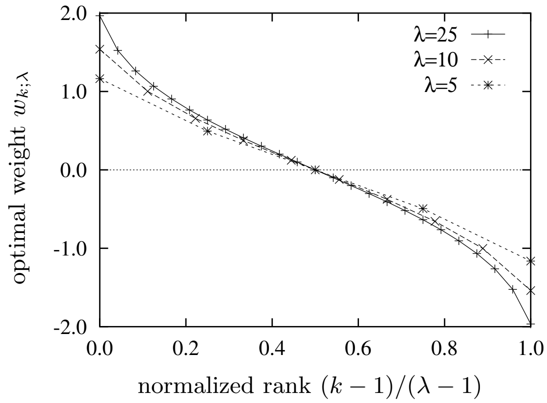

Fig. 2. Optimal weights $w_{k;\lambda} = E_{k;\lambda}$ plotted against the rank k of a candidate solution in the set of offspring for several values of λ. The ranks have been scaled linearly to fall in the range from zero to one. Note that only the points, not the connecting lines, are of practical significance.

Fig. 2 differ strongly from the step curves defined by Eq. (2) that describe the choice of weights characterizing the $(\mu/\mu, \lambda)$-ES. Finally, it is worth mentioning the relatively good correspondence between the left half of the curves in Fig. 2 and the (presumably empirically based) recommendations with regard to the choice of weights made in [14].

Inserting the optimal weights given in Eq. (13) into Eq. (12), it follows that the maximal quality gain of the $(\lambda)_{\text{opt}}$-ES is

$$\varDelta^* = \frac{1}{2}\sum_{k=1}^{\lambda} E_{k;\lambda}^2. \tag{14}$$

Defining $W_\lambda = \sum_{k=1}^{\lambda} E_{k;\lambda}^2$ and using results on properties of order statistics from [7], it can be seen that W_λ/λ asymptotically approaches unity as λ increases. Thus, the serial efficiency of the $(\lambda)_{\text{opt}}$-ES defined in Eq. (5) asymptotically approaches a value of 0.5, nearly two and a half times that of both the $(\mu/\mu, \lambda)$-ES and the $(1 + 1)$-ES. The dependence of the serial efficiency on the number of offspring generated per time step is illustrated in Fig. 3. It can be seen that the $(\lambda)_{\text{opt}}$-ES solidly outperforms not only the $(\mu/\mu, \lambda)$-ES, but it is also has a higher serial efficiency than EGS for all but the smallest values of λ.

4 Noise

As many real-world optimization problems are plagued by noise, the assumption that the fitness of a candidate solution can be determined exactly often is an idealization. In order to study the effects of noisy fitness evaluations on the performance of optimization strategies, it is frequently assumed that noise can be

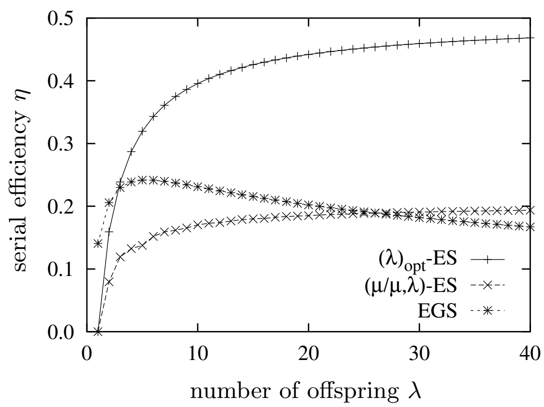

Fig. 3. Serial efficiency η of strategies on the sphere model in the limit $N \to \infty$ plotted against the number of offspring λ generated per time step. The curves represent results for the $(\lambda)_{\text{opt}}$-ES described by Eq. (14), the $(\mu/\mu, \lambda)$-ES analyzed in [11], and EGS studied in [2].

modeled by means of an additive, Gaussian term. That is, it is assumed the evaluation of a candidate solution $\mathbf{y}$ yields a value that is normally distributed with mean $f(\mathbf{y})$ and with variance ${\sigma_\epsilon^*}^2$, where σ_ϵ^* is referred to as the noise strength and may vary with the location in search space. See [1] for comprehensive results with regard to the effects of noise on various $(\mu/\rho \overset{+}{,} \lambda)$-ES.

An evolution strategy that has been found to be particularly robust with regard to the effects of noise is the $(\mu/\mu, \lambda)$-ES. In [3] it has been seen that this robustness is to be be attributed to the genetic repair effect. The term "genetic repair" has been introduced by Beyer [11] and refers to statistical error correction properties inherent in the multirecombination procedure. Typically, genetic repair affords the ability to operate with mutation strengths that increase (for the sphere model roughly linearly) with the number of candidate solutions generated per time step. The accompanying increase in quality gain is also roughly linear in λ, opening up the possibility of linear speed-up in a parallel implementation. In the presence of noise, the increased mutation strengths have been found to yield the additional benefit of reducing the noise-to-signal ratio $\vartheta = \sigma_\epsilon^*/\sigma^*$ that the strategy operates under. As seen in a comparison with other direct search strategies in [5], that benefit can be very substantial.

In the light of the results from the previous section, it seems interesting to ask whether the $(\lambda)_{\text{opt}}$-ES is capable of outperforming the $(\mu/\mu, \lambda)$-ES in the presence of noise as it does in its absence. At first sight, it appears that the $(\lambda)_{\text{opt}}$-ES is not able to benefit from genetic repair the way the $(\mu/\mu, \lambda)$-ES does. From Eq. (11) with the choice of weights $w_{k;\lambda} = E_{k;\lambda}$, it follows that the optimal normalized mutation strength of the $(\lambda)_{\text{opt}}$-ES in the absence of noise is unity and thus does not increase with increasing λ. The noise-to-signal ratio is unaffected. However, the system of equations (13) is solved not only by the above choice of weights, but also by the assignment

$$w_{k;\lambda} = \frac{E_{k;\lambda}}{\kappa}, \qquad k = 1, \ldots, \lambda, \tag{15}$$

for any $\kappa > 0$.[2] With this modified choice of weights, it follows from Eq. (12) that the optimal quality gain in the absence of noise according to Eq. (14) is unchanged. However, considering Eq. (11), it is clear that the mutation strength at which this quality gain is attained is κ and can thus be large if κ is chosen to be large.

The effect of the scaling of the weights in Eq. (15) is reminiscent of the use of rescaled mutations in the $(1, \lambda)$-ES proposed in [18] and analyzed in [9, 10]. The idea behind using rescaled mutations is to generate offspring using a high mutation strength, but to update the search point using a much smaller step length. A large mutation strength has the advantage of affording a strong signal component for selection that can outweigh any noise that is present, and to thus yield a good search direction. However, it is also likely to lead to a set of offspring all of which are inferior to the parent they are generated from. It is thus only the direction, not the length of the step that is used by the strategy. An evolution strategy using rescaled mutations updates the search point by using a progress vector that is shorter by a constant factor compared to the selected mutation vector.

It has been seen in [3] that the genetic repair effect resulting from multirecombination has the effect of providing an implicit rescaling. For the $(\mu/\mu, \lambda)$-ES, while the mutation vectors are of expected squared length N, the expectation of the squared length of the progress vector is N/μ. Similarly, for $(\lambda)_{\text{opt}}$-ES, the expected squared length of the mutation vectors is N, but with the choice of weights in Eq. (15), the expectation of the squared length of the progress vector is NW_λ/κ^2 according to Eq. (7). The choice of κ for the $(\lambda)_{\text{opt}}$-ES is thus similar to the choice of μ for the $(\mu/\mu, \lambda)$-ES in that it affords control over the amount of implicit rescaling inherent in the multirecombination process. Generally, larger values of κ can be expected to afford greater robustness in the presence of noise.

In order to derive a quality gain law for the $(\lambda)_{\text{opt}}$-ES in the presence of noise from Eq. (4), expected values of the overall squared length and of the signed length of the central component of the progress vector need to be computed in a fashion analogous to Sect. 3. Equation (7) for the squared length of the progress vector still holds as its derivation is unaffected by the presence of noise. The computation of the expected signed length of the progress vector's central component is less straightforward. For the purpose of selection, the candidate solutions are ordered according to their noisy fitness values. However, it is the *true* fitness values that determine the signed lengths of the central components of the respective mutation vectors. Technically, those signed lengths are concomitants of the order statistics. See [12] for an introduction to the topic and see [1, 3, 11] for the application to the problem of selection under Gaussian

[2] Negative κ also solve the system of equations, but correspond to extrema of the quality gain that are minima rather than maxima.

fitness noise. In the latter references it is shown that the probability density function of the concomitant $z_A^{(k;\lambda)}$ of the $(\lambda+1-k)$th order statistic is

$$p_{k;\lambda}(x) = \frac{1}{2\pi\vartheta}\frac{\lambda!}{(\lambda-k)!(k-1)!}\mathrm{e}^{-\frac{1}{2}x^2}\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}\left(\frac{y-x}{\vartheta}\right)^2\right) \left[\Phi\left(\frac{y}{\sqrt{1+\vartheta^2}}\right)\right]^{\lambda-k}\left[1-\Phi\left(\frac{y}{\sqrt{1+\vartheta^2}}\right)\right]^{k-1}\mathrm{d}y,$$

where $\vartheta = \sigma_\epsilon^*/\sigma^*$ denotes the noise-to-signal ratio that the strategy operates under, and where $\sigma_\epsilon^* = \sigma_\epsilon N/2R^2$ is the normalized noise strength. Using this density to replace Eq. (8), simple calculations analogous to those in [3] show that

$$\mathrm{E}\left[z_A^{(k;\lambda)}\right] = \frac{E_{k;\lambda}}{\sqrt{1+\vartheta^2}},$$

and therefore that in generalization of Eq. (9),

$$\mathrm{E}\left[z_A^{(\mathrm{avg})}\right] \overset{N\to\infty}{=} \frac{1}{\sqrt{1+(\sigma_\epsilon^*/\sigma^*)^2}}\sum_{k=1}^{\lambda} w_{k;\lambda}E_{k;\lambda}. \tag{16}$$

Thus, using Eqs. (7) and (16) in Eq. (4) and choosing the weights according to Eq. (15) it follows that the quality gain of the $(\lambda)_{\mathrm{opt}}$-ES on the sphere model in the presence of Gaussian noise is

$$\Delta^* \overset{N\to\infty}{=} \frac{W_\lambda}{\kappa}\left[\frac{{\sigma^*}^2}{\sqrt{{\sigma^*}^2+{\sigma_\epsilon^*}^2}} - \frac{{\sigma^*}^2}{2\kappa}\right]. \tag{17}$$

The dependence of the optimal mutation strength and of the resulting quality gain on the noise strength is illustrated in Fig. 4. The graphs look the same as the corresponding graphs for the $(\mu/\mu,\lambda)$-ES in [3] except for the different scaling of the axes. It can be inferred from the figures that while the $(\mu/\mu,\lambda)$-ES is capable of achieving positive quality gain up to a noise strength of $\sigma_\epsilon^* = 2\mu c_{\mu/\mu,\lambda}$, where $c_{\mu/\mu,\lambda}$ is the $(\mu/\mu,\lambda)$-ES progress coefficient defined in [11], the $(\lambda)_{\mathrm{opt}}$-ES does not need to stagnate up to a noise strength of $\sigma_\epsilon^* = 2\kappa$. It is important to note however that practically, finite search space dimensionalities set limits on the useful parameter values in both cases. As the degree of accuracy of the quality gain law of the $(\mu/\mu,\lambda)$-ES in [3] decreases with increasing μ and λ, so does that of the $(\lambda)_{\mathrm{opt}}$-ES in Eq. (17) when λ and κ are increased.

5 Cumulative Step Length Adaptation

In the considerations so far, the mutation strength has always been treated as an external parameter. In practice, of course, it needs to be adapted continually by the strategy, making the evolutionary algorithm together with the fitness environment it operates in a dynamic system. One mechanism for the adaptation

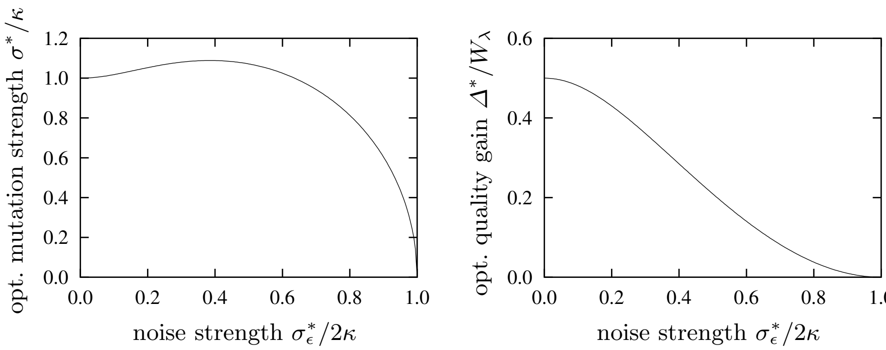

Fig. 4. Optimal mutation strength and corresponding quality gain of the $(\lambda)_{\text{opt}}$-ES plotted against the noise strength σ_ϵ^*. Due to the scaling of the axes, the curves are independent of the choice of λ and κ.

of the mutation strength is the cumulative step length adaptation procedure introduced by Ostermeier et al. [16]. In this section, that procedure is formulated for the $(\lambda)_{\text{opt}}$-ES with the choice of weights from Eq. (15) and then analyzed for the sphere model. The derivation remains sketchy in places as it closely parallels that for the $(\mu/\mu, \lambda)$-ES presented in [1, 6].

The goal of cumulative step length adaptation is to minimize correlations between successive steps. For that purpose, an exponentially fading record of the most recently taken steps is kept by accumulating progress vectors. Specifically, N-dimensional vector $\mathbf{s}$ is defined by $\mathbf{s}^{(0)} = \mathbf{0}$ and

$$\mathbf{s}^{(t+1)} = (1-c)\mathbf{s}^{(t)} + \kappa\sqrt{\frac{c(2-c)}{W_\lambda}}\mathbf{z}^{(\text{avg})}, \tag{18}$$

where superscripts indicate time steps. The cumulation parameter c is set to $1/\sqrt{N}$ according to a recommendation made in [13]. Recalling from Sect. 3 that the expected squared length of the progress vector $\mathbf{z}^{(\text{avg})}$ approximately equals NW_λ/κ^2, it is easy to verify that the choice of coefficients in Eq. (18) ensures that the distribution of the components of the accumulated progress vector $\mathbf{s}$ tends to standardized normality if the ordering of candidate solutions according to fitness values is random. It is also easily seen that Eq. (18) parallels the definition given in [14] for the case of weighted recombination. As in [6], the mutation strength is then adapted according to

$$\sigma^{(t+1)} = \sigma^{(t)} \exp\left(\frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right), \tag{19}$$

where the damping parameter D is set to $1/c$ as suggested in [13]. As a result of Eq. (19), the mutation strength is increased if the squared length of $\mathbf{s}$ exceeds N, which is a sign of positive correlations in the sequence of most recently taken steps. Conversely, the mutation strength is decreased if the squared length of $\mathbf{s}$ is less than N, which indicates negative correlations.

The analysis of cumulative step length adaptation presented in [1, 6] for the case of the $(\mu/\mu, \lambda)$-ES proceeds in three steps:

1. The accumulated progress vector $\mathbf{s}$ is decomposed into its central and lateral components, and recursive equations are derived for the overall squared length $\|\mathbf{s}\|^2$ and for the signed length s_A of the central component of that vector.
2. Expectations are taken and terms that become irrelevant in the limit $N \to \infty$ are dropped.
3. It is made use of the scale invariance properties of the quantities considered by demanding that their expected values do not change from one time step to the next.

The result of that procedure are two equations that can be used to determine expected values of $\|\mathbf{s}\|^2$ and s_A. We refrain from presenting detailed calculations for the case of the $(\lambda)_{\text{opt}}$-ES here as they are exactly analogous to those for the $(\mu/\mu, \lambda)$-ES. Also, as fluctuations of all quantities are neglected in the process, for the sake of brevity and readability, the equations are written in the quantities themselves rather than in their expected values. It is understood that for formal correctness, expectation operators need to be added throughout. The resulting equations read

$$\|\mathbf{s}\|^2 = (1-c)^2\|\mathbf{s}\|^2 + 2(1-c)\kappa\sqrt{\frac{c(2-c)}{W_\lambda}}s_A z_A^{(\text{avg})} + c(2-c)\frac{\kappa^2}{W_\lambda}\|\mathbf{z}^{(\text{avg})}\|^2 \tag{20}$$

and

$$s_A = (1-c)s_A + \kappa\sqrt{\frac{c(2-c)}{W_\lambda}}\left(z_A^{(\text{avg})} - \sigma^*\frac{\|\mathbf{z}^{(\text{avg})}\|^2}{N}\right). \tag{21}$$

They differ from their respective equivalents for the $(\mu/\mu, \lambda)$-ES derived in [1, 6] only in the coefficients. Using the relationships $z_A^{(\text{avg})} = W_\lambda/(\sqrt{1+\vartheta^2}\kappa)$ and $\|\mathbf{z}^{(\text{avg})}\|^2/N = W_\lambda/\kappa^2$ that follow from Eqs. (16) and (7) with Eq. (15), it follows that solving Eq. (21) for the expected signed length of the central component of the accumulated progress vector yields

$$s_A = \sqrt{\frac{W_\lambda(2-c)}{c}}\left(\frac{1}{\sqrt{1+\vartheta^2}} - \frac{\sigma^*}{\kappa}\right).$$

Inserting this result in Eq. (20) and rearranging terms yields

$$\|\mathbf{s}\|^2 = N + \frac{2(1-c)}{c}\frac{W_\lambda}{\sqrt{1+\vartheta^2}}\left(\frac{1}{\sqrt{1+\vartheta^2}} - \frac{\sigma^*}{\kappa}\right) \tag{22}$$

for the expected squared length of the accumulated progress vector.

The target mutation strength of the strategy is the mutation strength that cumulative step length adaptation seeks to attain and thus does not affect a change for. For the $(\lambda)_{\text{opt}}$-ES with cumulative step length adaptation, the target mutation strength is the mutation strength for which the argument to the

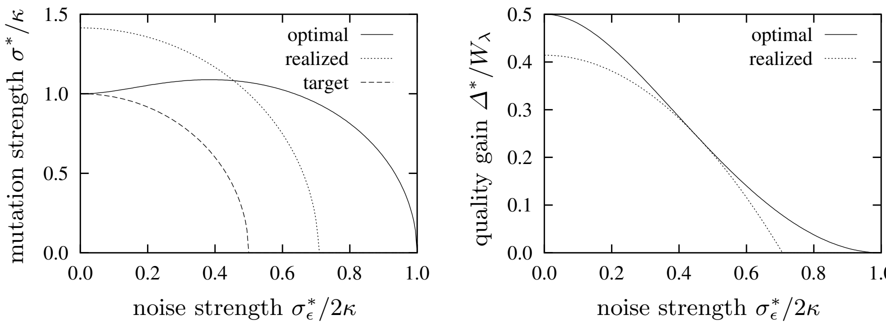

Fig. 5. Mutation strength σ^* and resulting quality gain Δ^* plotted against the noise strength σ_ϵ^*. In both graphs, the solid curves represent the optimal values from Fig. 4 and the dotted curves correspond to the values realized by the strategy and described by Eqs. (24) and (25), respectively. The dashed curve in the left hand graph is the target mutation strength given in Eq. (23).

exponential function in Eq. (19) equals zero. Using Eq. (22) and the fact that $\vartheta = \sigma_\epsilon^*/\sigma^*$, it follows that that mutation strength is

$$\sigma^* = \kappa\sqrt{1 - \left(\frac{\sigma_\epsilon^*}{\kappa}\right)^2}. \tag{23}$$

The dependence of the target mutation strength on the noise strength is illustrated and compared with the optimal mutation strength derived in Sect. 4 in the left hand graph of Fig. 5. While the shape of the curves is the same as in the corresponding graph for the $(\mu/\mu, \lambda)$-ES in [6], it is important to note that the scaling of the axes is different. However, in both cases, the target mutation strength agrees with the optimal mutation strength only in the case of no noise being present. For nonzero noise strengths, target mutation strengths are consistently below optimal mutation strengths. For the $(\mu/\mu, \lambda)$-ES it has been found in [6] that better agreement of the target mutation strength with the optimal mutation strength can be achieved by increasing μ and λ. For the $(\lambda)_{\text{opt}}$-ES, the same effect can be obtained by increasing κ with the purpose of moving closer to the left hand edge of the graph.

Finally, it is important to emphasize that as the $(\mu/\mu, \lambda)$-ES, the $(\lambda)_{\text{opt}}$-ES never in fact attains its target mutation strength. As adaptation is gradual rather than instantaneous, and as the distance to the optimizer continually decreases, the strategy will always be "behind" its target. Expanding the exponential function in Eq. (19) into a Taylor series, taking the decrease in distance to the optimizer into account, dropping all but the first terms, and demanding stationarity in the sense that the normalized mutation strength does not change yields equation

$$\sigma^* = \sigma^*\left(1 + \frac{\Delta^*}{N} + \frac{\|\mathbf{s}\|^2 - N}{2DN}\right).$$

For a more detailed derivation, see [1]. Inserting Eqs. (17) and (22) and solving for σ^*, it follows that the mutation strength actually realized by the strategy is

$$\sigma^* = \kappa\sqrt{2 - \left(\frac{\sigma_\epsilon^*}{\kappa}\right)^2} \tag{24}$$

if $\sigma_\epsilon^* \leq \sqrt{2}\kappa$, and it is zero if $\sigma_\epsilon^* > \sqrt{2}\kappa$. Inserting this result in Eq. (17) it follows that the resulting quality gain of the $(\lambda)_{\mathrm{opt}}$-ES with cumulative step length adaptation is

$$\Delta^* = \frac{\sqrt{2}-1}{2} W_\lambda \left(2 - \left(\frac{\sigma_\epsilon^*}{\kappa}\right)^2\right) \tag{25}$$

for $\sigma_\epsilon^* \leq \sqrt{2}\kappa$, and it is zero for $\sigma_\epsilon^* > \sqrt{2}\kappa$. Both the mutation strength and the corresponding quality gain are illustrated in Fig. 5. While the mutation strength realized by the strategy generally differs from the optimal mutation strength, the right hand graph shows that the loss in quality gain is quite acceptable provided that the strategy operates not too close to the right hand edge of the graphs. For the $(\mu/\mu, \lambda)$-ES, the recipe for achieving this is to increase μ and λ; for the $(\lambda)_{\mathrm{opt}}$-ES, it is to increase κ.

6 Summary and Conclusions

In this paper, the behavior of weighted multirecombination evolution strategies has been studied on the infinite-dimensional sphere model. Optimal rank-based weights have been computed, and it has been found that optimal performance is achieved if those weights are set to equal the expected values of the order statistics of the standardized normal distribution. The performance of the resulting strategy — referred to as $(\lambda)_{\mathrm{opt}}$-ES — has been analyzed, and it was seen that unlike the $(\mu/\mu, \lambda)$-ES, the $(\lambda)_{\mathrm{opt}}$-ES is capable of exceeding the serial efficiency of the $(1+1)$-ES by a factor of roughly two and a half. It has then been found that the $(\lambda)_{\mathrm{opt}}$-ES in its original form does not benefit from genetic repair in the sense that a larger number of offspring generated per time step allows it to operate with larger mutation strengths. However, the strategy can be modified by scaling all weights using a common factor κ. While that factor is without influence on the performance of the strategy if there is no noise present, it has been found to be able to contribute positively to the strategy's robustness in the presence of Gaussian fitness noise. The scaling of weights has been likened to the idea of using rescaled mutations to which it is similar in effect, but from which it differs in that no explicit rescaling is required. Rather, the possibility of making large trial steps and at the same time small search steps is an implicit result of weighted multirecombination in combination with an appropriate choice of weights. Finally, it has been seen that by virtue of a simple modification, the cumulative step length adaptation mechanism works for the $(\lambda)_{\mathrm{opt}}$-ES as well as it does for the $(\mu/\mu, \lambda)$-ES, and that good mutation strength settings can

Table 1. Comparison of properties of $(\mu/\mu,\lambda)$-ES and $(\lambda)_{\text{opt}}$-ES on the sphere model.

	$(\mu/\mu,\lambda)$-ES	$(\lambda)_{\text{opt}}$-ES
quality gain Δ^*	$\dfrac{\sigma^{*2}c_{\mu/\mu,\lambda}}{\sqrt{\sigma^{*2}+\sigma_\epsilon^{*2}}}-\dfrac{\sigma^{*2}}{2\mu}$	$\dfrac{W_\lambda}{\kappa}\left(\dfrac{\sigma^{*2}}{\sqrt{\sigma^{*2}+\sigma_\epsilon^{*2}}}-\dfrac{\sigma^{*2}}{2\kappa}\right)$
optimal σ^* (w/o noise)	$\mu c_{\mu/\mu,\lambda}$	κ
optimal Δ^* (w/o noise)	$\mu c_{\mu/\mu,\lambda}^2/2$ ($\stackrel{\lambda\to\infty}{\longrightarrow}$ 0.202λ)	$W_\lambda/2$ ($\stackrel{\lambda\to\infty}{\longrightarrow}$ 0.5λ)
maximal σ_ϵ^*	$2\mu c_{\mu/\mu,\lambda}$	2κ

typically be arrived at by choosing κ sufficiently large. Table 1 summarizes some of the most important findings with regard to the performance of the $(\lambda)_{\text{opt}}$-ES on the sphere model and constrasts them with the corresponding results for the $(\mu/\mu,\lambda)$-ES.

Finally, it is important to emphasize that all results in this paper have been derived under the assumption of infinite search space dimensionality. The findings help to provide a good intuitive understanding of the influence of the parameters λ and κ, of the issues involved in the choice of weights, and of the consequences of that choice for genetic repair and the performance of multi-recombination evolution strategies. However, for the case of the $(\mu/\mu,\lambda)$-ES it has been seen that the accuracy of the predictions made on the basis of the assumption of infinite search space dimensionality is not very good in finite-dimensional search spaces unless N is very large and λ is rather small. An improved approximation for the noisy case was derived in [4] that takes into account fluctuations in the squared length of the mutation vectors, and that was used as a basis for computing optimal values of the population size parameters μ and λ. A similar investigation for the $(\lambda)_{\text{opt}}$-ES remains to be conducted with the goal of determining optimal settings for λ and κ, and for verifying to what degree the performance advantages predicted can be observed in finite-dimensional search spaces.

Acknowledgments

The author is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). This paper was born out of discussions at the 2004 Dagstuhl Seminar on the Theory of Evolutionary Algorithms.

References

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Genetic Algorithms and Evolutionary Computation Series. Kluwer Academic Publishers, Boston, 2002.

[2] D. V. Arnold. An analysis of evolutionary gradient search. In *Proc. of the 2004 IEEE Congress on Evolutionary Computation*, pages 47–54. IEEE Press, Piscataway, NJ, 2004.

[3] D. V. Arnold and H.-G. Beyer. Local performance of the $(\mu/\mu_I, \lambda)$-ES in a noisy environment. In W. N. Martin and W. M. Spears, editors, *Foundations of Genetic Algorithms 6*, pages 127–141. Morgan Kaufmann Publishers, San Francisco, 2001.
[4] D. V. Arnold and H.-G. Beyer. Performance analysis of evolution strategies with multi-recombination in high-dimensional $I\!R^N$-search spaces disturbed by noise. *Theoretical Computer Science*, 289(1):629–647, 2002.
[5] D. V. Arnold and H.-G. Beyer. A comparison of evolution strategies with other direct search methods in the presence of noise. *Computational Optimization and Applications*, 24(1):135–159, 2003.
[6] D. V. Arnold and H.-G. Beyer. Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49(4):617–622, 2004.
[7] N. Balakrishnan and C. R. Rao. Order statistics: An introduction. In N. Balakrishnan and C. R. Rao, editors, *Handbook of Statistics*, volume 16, pages 3–24. Elsevier, Amsterdam, 1998.
[8] H.-G. Beyer. On the asymptotic behavior of multirecombinant evolution strategies. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature — PPSN IV*, pages 122–133. Springer Verlag, Heidelberg, 1996.
[9] H.-G. Beyer. Mutate large, but inherit small! On the analysis of rescaled mutations in $(\tilde{1}, \tilde{\lambda})$-ES with noisy fitness data. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature — PPSN V*, pages 109–118. Springer Verlag, Heidelberg, 1998.
[10] H.-G. Beyer. Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *Computer Methods in Mechanics and Applied Engineering*, 186:239–267, 2000.
[11] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer Verlag, Heidelberg, 2001.
[12] H. A. David and H. N. Nagaraja. Concomitants of order statistics. In N. Balakrishnan and C. R. Rao, editors, *Handbook of Statistics*, volume 16, pages 487–513. Elsevier, Amsterdam, 1998.
[13] N. Hansen. *Verallgemeinerte individuelle Schrittweitenregelung in der Evolutionsstrategie*. Mensch & Buch Verlag, Berlin, 1998.
[14] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
[15] S. Kern, S. D. Müller, N. Hansen, D. Büche, J. Ocenasek, and P. Koumoutsakos. Learning probability distributions in continuous evolutionary algorithms — A comparative review. *Natural Computing*, 3(1):77–112, 2004.
[16] A. Ostermeier, A. Gawelczyk, and N. Hansen. Step-size adaptation based on non-local use of selection information. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature — PPSN III*, pages 189–198. Springer Verlag, Heidelberg, 1994.
[17] I. Rechenberg. *Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Friedrich Frommann Verlag, Stuttgart, 1973.
[18] I. Rechenberg. *Evolutionsstrategie '94*. Friedrich Frommann Verlag, Stuttgart, 1994.
[19] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Verlag Dr. Kovač, Hamburg, 1997.
[20] R. Salomon. Evolutionary search and gradient search: Similarities and differences. *IEEE Transactions on Evolutionary Computation*, 2(2):45–55, 1998.