On the Effects of Outliers on Evolutionary Optimization

Dirk V. Arnold and Hans-Georg Beyer

Department of Computer Science XI
University of Dortmund
44221 Dortmund, Germany
{dirk.arnold,hans-georg.beyer}@cs.uni-dortmund.de

Abstract. Most studies concerned with the effects of noise on evolutionary computation have assumed a Gaussian noise model. However, practical optimization strategies frequently face situations where the noise is not Gaussian, and sometimes it does not even have a finite variance. In particular, outliers may be present. In this paper, Cauchy distributed noise is used for modeling such situations. A performance law that describes how the progress of an evolution strategy using intermediate recombination scales in the presence of such noise is derived. Implications of that law are studied numerically, and comparisons with the case of Gaussian noise are drawn.

1 Introduction

In studies of optimization strategies, it is frequently assumed that the objective function value of a candidate solution can be determined exactly. However, given the noisy nature of many real-world optimization problems, that assumption often is an idealization. Noise can stem from sources as different as — and not restricted to — measurement limitations, the use of randomized algorithms, incomplete sampling of large spaces, and human computer interaction. Understanding how noise impacts the performance of optimization strategies is important for choosing appropriate strategy variants, for the sizing of strategy parameters, and for the design of new, more noise resistant algorithms.

A number of studies have dealt with the effects of noise on the performance of genetic algorithms. Fitzpatrick and Grefenstette [11] have explored the tradeoff between averaging over multiple fitness evaluations versus increasing the population size. Rattray and Shapiro [13] have studied finite population effects in the presence of additive Gaussian noise. Miller and Goldberg [12] have investigated the effect of Gaussian noise on different selection mechanisms. A more extensive overview of related work can be found in [8].

In the realm of evolution strategies, we have studied the effects of noise on the local performance of the algorithms by considering a noisy version of the sphere model. A comprehensive summary of the work can be found in [2]. In [5], the performance of the $(1+1)$-ES has been studied. It was found that the overvaluation of the fitness of candidate solutions that results from the use of plus-selection severely affects both the local performance of the strategy and the functioning of success probability based step length adaptation mechanisms. In [4], a performance law has been derived for the $(\mu/\mu, \lambda)$-ES with intermediate recombination. In [3], that performance law has been used to address the issue of resampling in order to reduce the amount of noise present. It was found that in contrast to results obtained for the $(1, \lambda)$-ES, for the $(\mu/\mu, \lambda)$-ES, increasing the population size is preferable to averaging over multiple samples if the truncation ratio μ/λ is chosen appropriately. This is an encouraging result as it shows that the evolution strategy is able to handle the noise more effectively than by blind averaging. The influence of finite search space dimensionalities has been explored in [6]. Finally, in [7], the performance of evolution strategies in the presence of noise has been compared with that of other direct search algorithms.

All of the aforementioned studies as well as many other investigations of the effects of noise on the performance of optimization strategies have in common that either Gaussian noise or at least noise of a finite variance is assumed. However, it is doubtful whether Gaussian noise satisfactorily models all variants of noise that occur in practical applications. There is a possibility that the results that have been obtained may qualitatively depend on that assumption.

Beyer, Olhofer, and Sendhoff [9] have considered particular situations involving non-Gaussian noise. Their approach is to apply transformations that make the noise nearly Gaussian and to then use the results obtained for the case of Gaussian noise. While proceeding as such extends significantly the realm of situations that can be considered, there are situations where a transformation that makes the noise nearly Gaussian is not possible. In particular, in practice, optimization strategies frequently face outliers. In order to model such situations, noise distributions with tails much longer than those of a normal distribution need to be considered. One such distribution is the Cauchy distribution that has impacted the optimization literature in the past not as a noise model but as a mutation strategy. Szu and Hartley [17] have suggested to use Cauchy distributed mutations in simulated annealing in order to better be able to escape local optima by occasional long jumps. Rudolph [15] has studied the effects of Cauchy distributed mutations in evolution strategy optimization.

In the present paper, we investigate the effects that outliers have on the performance of the $(\mu/\mu, \lambda)$-ES by considering Cauchy distributed noise. The choice of strategy is motivated both by the fact that it is relatively amenable to mathematical analysis and by its proven good performance. In Sect. 2, the strategy as well as the fitness environment considered are introduced. As outliers need to be modeled, a transformation to normality of the noise is not possible. Also, as the Cauchy distribution does not have finite moments, an approach using expansions of probability distributions in terms of their moments is excluded. In Sect. 3, the expected average of concomitants of selected Cauchy order statistics is computed. As a result of the calculations, numerical comparisons of the effects of Cauchy noise with those of Gaussian noise can be performed in Sect. 4. We conclude with a brief discussion of the results and suggest directions for future research.

2 Preliminaries

The $(\mu/\mu, \lambda)$-ES in every time step generates $\lambda > \mu$ offspring candidate solutions from a population of μ parents and subsequently replaces the parental population by the μ best of the offspring. Using isotropic normal mutations, for real-valued objective functions $f : \mathbb{R}^N \to \mathbb{R}$, generation of an offspring candidate solution consists of adding a vector $\sigma\mathbf{z}$, where $\mathbf{z}$ consists of independent, standard normally distributed components, to the centroid of the parental population. The standard deviation σ of the components of vector $\sigma\mathbf{z}$ is referred to as the mutation strength, vector $\mathbf{z}$ as a mutation vector. The average of those mutation vectors that correspond to offspring candidate solutions that are selected to form the population of the next time step is the progress vector $\langle\mathbf{z}\rangle$.

Since the early work of Rechenberg [14], the local performance of evolution strategies has commonly been studied on a class of functions known as the sphere model. The sphere model is the set of all functions $f : \mathbb{R}^N \to \mathbb{R}$ with

$$f(\mathbf{x}) = g(\|\hat{\mathbf{x}} - \mathbf{x}\|),$$

where $g : \mathbb{R} \to \mathbb{R}$ is a strictly monotonic function of the distance $R = \|\mathbf{R}\| = \|\hat{\mathbf{x}} - \mathbf{x}\|$ of a candidate solution $\mathbf{x}$ from the optimizer $\hat{\mathbf{x}}$. It has frequently served as a model for fitness landscapes at a stage where the population of candidate solutions is in relatively close proximity to the optimizer and is most often studied in the limit of very high search space dimensionality. In this paper, it is assumed that there is noise present in the process of evaluating the objective function in that evaluating a candidate solution $\mathbf{x}$ does not yield the candidate solution's true fitness $f(\mathbf{x})$, but a noisy fitness

$$f_\epsilon(\mathbf{x}) = f(\mathbf{x}) + \sigma_\epsilon Z, \tag{1}$$

where Z is a random variable. While in our previous work, we had always assumed that the distribution of Z is standard normal, in the present paper, the case that Z is drawn from a Cauchy distribution is investigated.

A commonly used measure for the performance of evolution strategies on the sphere model is the progress rate

$$\varphi^{(t)} = \mathrm{E}\left[\langle R\rangle^{(t)} - \langle R\rangle^{(t+1)}\right]$$

that is defined as the expectation of the decrease in the distance between the population centroid and the optimizer in a single time step. The commonly used approach to computing the progress rate relies on a decomposition of vectors 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 $\mathbf{R} = \hat{\mathbf{x}} - \mathbf{x}$ and $\mathbf{z}_B$ is in the hyperplane perpendicular to that. In the present context, $\mathbf{z}$ can be either a mutation vector or a progress vector. 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.

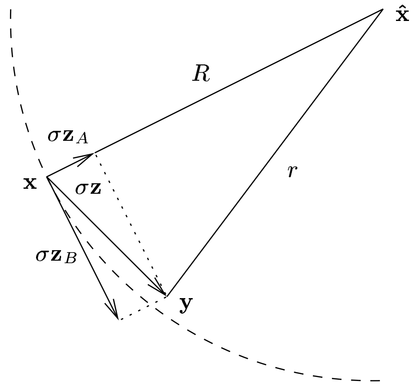

Fig. 1. *Decomposition of a vector $\mathbf{z}$ into central component $\mathbf{z}_A$ and lateral component $\mathbf{z}_B$. Vector $\mathbf{z}_A$ is parallel to $\hat{\mathbf{x}} - \mathbf{x}$, vector $\mathbf{z}_B$ is in the hyperplane perpendicular to that. The starting and end points, $\mathbf{x}$ 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.*

In what follows, we make use of a number of simplifications that hold exactly in the limit $N \to \infty$, but that have been seen to provide good approximations for moderately large values of N already. A more complete justification of the argument that follows and that remains somewhat sketchy due to space limitations can be found in [4].

The $(\mu/\mu, \lambda)$-ES applies all mutations to a single point — the population centroid. Selection of candidate solutions is on the basis of their (noisy) fitness. The lateral components of all mutation vectors contribute equally to the fitness of the offspring candidate solutions they generate. They are thus selectively neutral. The contribution of the central components of the mutation vectors to the fitness of the offspring candidate solutions is asymptotically normal. Introducing normalizations

$$\sigma^* = \sigma \frac{N}{\langle R \rangle}, \quad \sigma^*_\epsilon = \sigma_\epsilon \frac{N}{\langle R \rangle g'(\langle R \rangle)}, \quad \text{and} \quad \varphi^* = \varphi \frac{N}{\langle R \rangle},$$

and defining the noise-to-signal ratio $\vartheta = \sigma^*_\epsilon / \sigma^*$, the performance law

$$\varphi^* = \sigma^* M_{\mu/\mu,\lambda}(\vartheta) - \frac{{\sigma^*}^2}{2\mu} \tag{2}$$

has been derived in [4]. The first term on the right hand side of the equation is due to the central component of the progress vector, the second term is due to its lateral component. The factor μ in the denominator of the second term is a result of the independence of the lateral components of the mutation vectors and signifies the presence of genetic repair. The term $M_{\mu/\mu,\lambda}(\vartheta)$ is frequently referred to as the progress coefficient and results from the (noisy) selection of μ of the λ offspring candidate solutions that have been generated. It is to be defined more formally and computed for the case of Cauchy noise in the next section.

3 Determining the Progress Coefficient

Let Y be a standard normally distributed random variable corresponding to the standardized contributions of the central components of the mutation vectors to the true fitness of the offspring candidate solutions they generate. Letting Z be the noise variate from Eq. (1), random variable X is defined as

$$X = Y + \vartheta Z \tag{3}$$

and reflects the standardized noisy fitness of an offspring candidate solution. We generate λ bivariate observations $(X_1, Y_1), \ldots, (X_\lambda, Y_\lambda)$ by λ times independently sampling Y and Z and using Eq. (3). We then order the observations by their X variates. The order statistics of X are denoted as usual by $X_{i:\lambda}$, $1 \leq i \leq \lambda$. That is, $X_{1:\lambda} \leq X_{2:\lambda} \leq \cdots \leq X_{\lambda:\lambda}$. The Y variate associated with $X_{i:\lambda}$ is called the concomitant of the ith order statistic and is denoted by $Y_{i;\lambda}$. The term $M_{\mu/\mu,\lambda}(\vartheta)$ from Sect. 2 is the expectation of the average of the concomitants of the μ largest order statistics

$$M_{\mu/\mu,\lambda}(\vartheta) = \mathrm{E}\left[\frac{1}{\mu}\sum_{i=1}^{\mu} Y_{\lambda-i+1;\lambda}\right]. \tag{4}$$

For the case of Gaussian noise, $M_{\mu/\mu,\lambda}(\vartheta)$ has been computed in [4]. The derivation that follows assumes a general noise distribution and closely parallels the aforementioned one. The specialization to the case of Cauchy noise will be presented in Sect. 4. The reader not interested in the particulars of the calculations may safely skip the following paragraphs and jump to Eq. (6) for the result.

Let $P(y) = \Phi(y)$ denote the cumulative distribution function (cdf) of the standardized normal distribution, and let $p(y) = \mathrm{d}\Phi/\mathrm{d}y = \exp(-y^2/2)/\sqrt{2\pi}$ denote the corresponding probability density function (pdf). Furthermore, let $P_\epsilon(z)$ and $p_\epsilon(z)$ denote the cdf and the pdf, respectively, of the ϑZ variate. The distribution of the X variate is the convolution of those of the other two variates and thus has pdf

$$q(x) = \int_{-\infty}^{\infty} p(y) p_\epsilon(x-y)\mathrm{d}y. \tag{5}$$

The corresponding cdf $Q(x)$ can be obtained by integration.

According to David and Nagaraja [10], the pdf of the concomitant of the ith order statistic is

$$p_{i;\lambda}(y) = \frac{\lambda!}{(\lambda-i)!(i-1)!} p(y) \int_{-\infty}^{\infty} p_\epsilon(x-y)[1-Q(x)]^{\lambda-i}[Q(x)]^{i-1}\mathrm{d}x.$$

Using this pdf in Eq. (4) and swapping the order of the integration and summation, the expectation of the average of the μ selected concomitants is

$$\begin{aligned} M_{\mu/\mu,\lambda}(\vartheta) &= \frac{1}{\mu}\sum_{i=1}^{\mu}\int_{-\infty}^{\infty} y p_{\lambda-i+1;\lambda}(y)\mathrm{d}y \\ &= \frac{\lambda!}{\mu}\int_{-\infty}^{\infty} y p(y) \int_{-\infty}^{\infty} p_\epsilon(x-y) \sum_{i=1}^{\mu} \frac{[Q(x)]^{\lambda-i}[1-Q(x)]^{i-1}}{(\lambda-i)!(i-1)!}\mathrm{d}x\mathrm{d}y. \end{aligned}$$

Using the identity (compare Abramowitz and Stegun [1], Eqs. **6.6.4** and **26.5.1**)

$$\sum_{i=1}^{\mu} \frac{Q^{\lambda-i}[1-Q]^{i-1}}{(\lambda-i)!(i-1)!} = \frac{1}{(\lambda-\mu-1)!(\mu-1)!}\int_0^Q z^{\lambda-\mu-1}[1-z]^{\mu-1}\mathrm{d}z,$$

it follows

$$M_{\mu/\mu,\lambda}(\vartheta) = (\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty} yp(y)\int_{-\infty}^{\infty} p_\epsilon(x-y) \int_0^{Q(x)} z^{\lambda-\mu-1}[1-z]^{\mu-1}\mathrm{d}z\mathrm{d}x\mathrm{d}y.$$

Substituting $z = Q(w)$ yields

$$M_{\mu/\mu,\lambda}(\vartheta) = (\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty} yp(y)\int_{-\infty}^{\infty} p_\epsilon(x-y) \int_{-\infty}^{x} q(w)[Q(w)]^{\lambda-\mu-1}[1-Q(w)]^{\mu-1}\mathrm{d}w\mathrm{d}x\mathrm{d}y.$$

Changing the order of the integrations results in

$$M_{\mu/\mu,\lambda}(\vartheta) = (\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty} q(w)[Q(w)]^{\lambda-\mu-1}[1-Q(w)]^{\mu-1}I(w)\mathrm{d}w,$$

where, using the fact that the mean of the standardized normal distribution is zero,

$$\begin{aligned} I(w) &= \int_{-\infty}^{\infty} yp(y)\int_w^{\infty} p_\epsilon(x-y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^{\infty} yp(y)[1-P_\epsilon(w-y)]\mathrm{d}y \\ &= \int_{-\infty}^{\infty} [-yp(y)]P_\epsilon(w-y)\mathrm{d}y. \end{aligned}$$

As $\mathrm{d}p/\mathrm{d}y = -yp(y)$, partial integration yields

$$I(w) = p(y)P_\epsilon(w-y)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} p(y)p_\epsilon(w-y)\mathrm{d}y.$$

The first of the two terms on the right hand side equals zero. Comparison of the second term with Eq. (5) shows that $I(w) = q(w)$ and therefore that

$$M_{\mu/\mu,\lambda}(\vartheta) = (\lambda-\mu)\binom{\lambda}{\mu}\int_{-\infty}^{\infty} [q(w)]^2[Q(w)]^{\lambda-\mu-1}[1-Q(w)]^{\mu-1}\mathrm{d}w. \qquad (6)$$

The remaining integral generally cannot be solved in closed form but needs to be evaluated numerically.

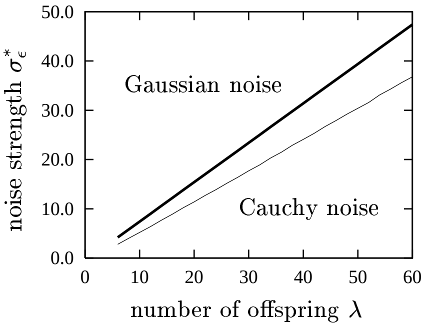

Fig. 2. *The normalized noise strength σ_ϵ^* up to which nonzero progress is possible as a function of the number of offspring λ generated per time step. The truncation ratio μ/λ is chosen to be optimal. The bold line represents the result for Gaussian noise, the thin line that for Cauchy noise.*

4 Gaussian Noise Versus Cauchy Noise

In case the distribution of the noise is Gaussian, the distribution of the X variates is the convolution of two normal variates and is thus itself normal. More specifically, the cdf of the X variate is $Q(w) = \Phi(w/\sqrt{1+\vartheta^2})$. Using the substitution $x = w/\sqrt{1+\vartheta^2}$ in Eq. (6), it follows

$$M_{\mu/\mu,\lambda}^{(\mathrm{Gauss})}(\vartheta) = \frac{c_{\mu/\mu,\lambda}}{\sqrt{1+\vartheta^2}} \tag{7}$$

for the progress coefficient, where

$$c_{\mu/\mu,\lambda} = \frac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}[\Phi(x)]^{\lambda-\mu-1}[1-\Phi(x)]^{\mu-1}\mathrm{d}x$$

is independent of the noise level ϑ and depends on the population size parameters μ and λ only. The result agrees with that from [4]. In that reference, it has been seen that the strong performance of the $(\mu/\mu,\lambda)$-ES in the presence of Gaussian noise is due to the presence of genetic repair. The factor μ in the denominator of the fraction in Eq. (2) not only reduces the term that makes a negative contribution to the progress rate, but it also has the effect that the search space can be explored at higher mutation strengths. Those increased mutation strengths decrease the noise-to-signal ratio $\vartheta = \sigma_\epsilon^*/\sigma^*$ that the strategy operates under and make the $(\mu/\mu,\lambda)$-ES vastly more efficient than other types of evolution strategy on the noisy sphere.

In the case of Cauchy noise, separating the influence of ϑ from that of μ and λ is not possible. The cdf and the pdf of the noise term are $P_\epsilon(z) = 1/2 + \arctan(z/\vartheta)/\pi$ and $p_\epsilon(z) = \vartheta/(\pi(\vartheta^2+z^2))$, respectively. In order to numerically evaluate the integral in Eq. (6), both the pdf and the cdf of the X variate need to be determined. For that purpose, either Eq. (5) can be used or an approach based on characteristic functions can be employed. The characteristic function of the Y variates is $\phi(t) = \exp(-t^2/2)$, that of the noise term in Eq. (3) is $\exp(-\vartheta|t|)$. The characteristic function of the convolution is the product of the two. According to Stuart and Ord [16], the pdf of the X variates can be obtained from that characteristic function and be written as

$$q(x) = \frac{1}{\pi}\mathrm{e}^{\vartheta^2/2}\int_0^\infty \cos(tx)\exp\left(-\frac{1}{2}(t+\vartheta)^2\right)\mathrm{d}t.$$

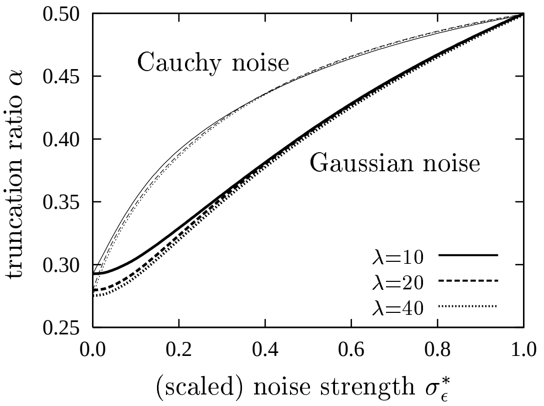

Fig. 3. *The optimal truncation ratio as a function of the normalized noise strength* σ_ϵ^*. *The normalized noise strengths on the horizontal axis have been scaled by division by the maximum normalized noise strength for which nonzero progress is possible. The bold lines represent results for Gaussian noise, the thin lines those for Cauchy noise.*

The corresponding cdf can be obtained by integration and be written as

$$Q(x) = \frac{1}{2} + \frac{1}{\pi} \mathrm{e}^{\vartheta^2/2} \int_0^\infty \frac{\sin(tx)}{t} \exp\left(-\frac{1}{2}(t+\vartheta)^2\right) \mathrm{d}t.$$

Using those relationships in the numerical evaluation of Eq. (6) yields results that make it possible to compare the performance of the $(\mu/\mu, \lambda)$-ES in the presence of Cauchy noise with that in the presence of Gaussian noise. It is important to keep in mind, however, that a naive quantitative comparison is problematic due to the basic incomparability of the parameter ϑ for the two types of distribution. While in the case of Gaussian noise ϑ is the standard deviation, in the case of Cauchy noise ϑ is simply a scale parameter the choice of which is somewhat arbitrary. Nonetheless, keeping that caveat in mind and proceeding with care, meaningful comparisons can be made.

One of the main results of the analysis in [4] is that in the presence of Gaussian noise, the $(\mu/\mu, \lambda)$-ES is capable of nonzero progress up to a normalized noise strength σ_ϵ^* that is proportional to the number of offspring λ generated per time step. As a consequence, by sufficiently increasing the population size, positive progress on the noisy sphere can be achieved for any noise strength. Increased population sizes make it possible to use larger mutation strengths that in turn reduce the noise-to-signal ratio ϑ. Figure 2 has been obtained by numerically evaluating Eq. (6) and suggests that that benefit of genetic repair is enjoyed also in the presence of Cauchy noise. As for Gaussian noise, the noise strength up to which progress is possible appears to be linear in λ.

A second important insight gained in [3] is that the $(\mu/\mu, \lambda)$-ES can partially compensate for a lack of reliable information by using the noisy information provided by a larger number of parents than it would optimally use in the absence of noise. While in the absence of noise, the $(\mu/\mu, \lambda)$-ES ideally operates

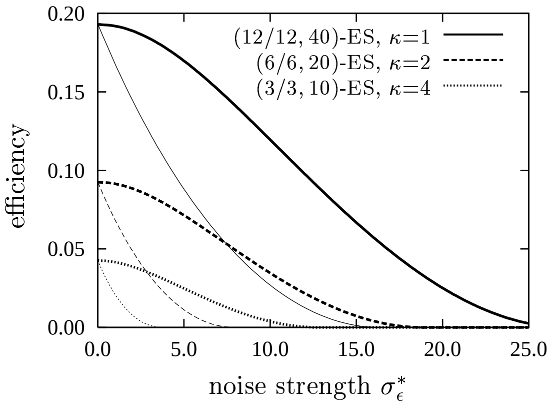

Fig. 4. *The efficiencies (defined as the normalized maximum progress rate per evaluation of the objective function) of various $(\mu/\mu, \lambda)$-ES that average κ objective function measurements as functions of the normalized noise strength σ_ϵ^*. The bold lines represent results for Gaussian noise, the thin lines those for Cauchy noise. Note that direct comparisons between curves corresponding to different noise models are relatively meaningless as explained in the text.*

with a truncation ratio $\alpha = \mu/\lambda$ in the vicinity of 0.27, that value gradually increases to 0.5 at the point where nonzero progress ceases to be possible. Figure 3 demonstrates that the same is true in the presence of Cauchy noise.

Finally, it has been seen in [3] that in the presence of Gaussian noise, larger population sizes are preferable to averaging over multiple evaluations of candidate solutions. The averaging in search space that is implicit in the recombination procedure of the $(\mu/\mu, \lambda)$-ES is more powerful than the explicit averaging of $\kappa > 1$ objective function values in fitness space. In the presence of Gaussian noise, averaging κ samples reduces the variance of the noise by a factor of κ. For Cauchy noise, the average of κ independent samples has the same distribution as the individual samples, rendering resampling entirely useless. Figure 4 confirms that by showing that while all of the strategies in that figure use the same number of objective function evaluations per time step, those that do not resample but rather rely on larger population sizes have the highest efficiencies.

To conclude, the differences between the effects that Gaussian noise and Cauchy noise have on the performance of the $(\mu/\mu, \lambda)$-ES are merely quantitative. Outliers that are frequent in the case of Cauchy noise do not lead to qualitatively new effects. The important conclusions with respect to the choice of population size parameters and to the use of resampling that were drawn for the case of Gaussian noise in previous studies remain valid in the case of Cauchy noise. It is important, however, to note that that result cannot be expected to hold for strategies in which candidate solutions can survive for several time steps and are not reevaluated periodically. The investigation of the behavior of such strategies as well as the analysis of the influence of finite search space dimensionalities along the lines of [6] remain as challenges for future research.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant Be 1578/6-3 as well as through the Collaborative Research Center SFB531.

References

1. M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1965.
2. D. V. Arnold. *Noisy Optimization with Evolution Strategies.* Genetic Algorithms and Evolutionary Computation Series. Kluwer Academic Publishers, Norwell, MA, 2002.
3. D. V. Arnold and H.-G. Beyer. Efficiency and self-adaptation of the $(\mu/\mu_I, \lambda)$-ES in a noisy environment. In M. Schoenauer et al., editor, *Parallel Problem Solving from Nature — PPSN VI*, pages 39–48. Springer Verlag, Berlin, 2000.
4. 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, San Francisco, 2001.
5. D. V. Arnold and H.-G. Beyer. Local performance of the $(1+1)$-ES in a noisy environment. *IEEE Transactions on Evolutionary Computation*, 6(1):30–41, 2002.
6. D. V. Arnold and H.-G. Beyer. Performance analysis of evolution strategies with multi-recombination in high-dimensional $\mathbb{R}^N$-search spaces disturbed by noise. *Theoretical Computer Science*, 289(1):629–647, 2002.
7. 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.
8. 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.
9. H.-G. Beyer, M. Olhofer, and B. Sendhoff. On the behavior of $(\mu/\mu_I, \lambda)$-ES optimizing functions disturbed by generalized noise. In J. Rowe et al., editor, *Foundations of Genetic Algorithms 7*. Morgan Kaufmann, San Francisco, 2003. In press.
10. 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.
11. J. M. Fitzpatrick and J. J. Grefenstette. Genetic algorithms in noisy environments. *Machine Learning*, 3:101–120, 1988.
12. B. L. Miller and D. E. Goldberg. Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, 4(2):113–131, 1997.
13. M. Rattray and J. Shapiro. Noisy fitness evaluation in genetic algorithms and the dynamics of learning. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms 4*, pages 117–139. Morgan Kaufmann, San Francisco, 1997.
14. I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach den Prinzipien der biologischen Evolution.* Frommann-Holzboog, Stuttgart, 1973.
15. G. Rudolph. Local convergence rates of simple evolutionary algorithms with Cauchy mutations. *IEEE Transactions on Evolutionary Computation*, 1(4):249–258, 1997.
16. A. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics. Volume 1: Distribution Theory.* Edward Arnold, London, sixth edition, 1994.
17. H. Szu and R. Hartley. Fast simulated annealing. *Physics Letters A*, 122:157–162, 1987.