FHV FORSCHUNG

Forschungszentrum
Prozess- und Produkt-Engineering

ARBEITSBERICHT PROZESS- UND PRODUKT-ENGINEERING

Technical Report 2015/02

Population Size Control of CMSA-ES for Noisy Optimization Using Time Series Analysis

FH VORARLBERG

Michael Hellwig
Michael.Hellwig@fhv.at

Hans-Georg Beyer
Hans-Georg.Beyer@fhv.at

30. Juni 2015

Population Size Control of CMSA-ES for Noisy Optimization Using Time Series Analysis

Michael Hellwig and Hans-Georg Beyer

June 30, 2015

1 Introduction

For many real-world applications the problem complexity is increased by noise perturbations. The noise may stem from many different sources such as randomized simulations or sensory disturbances. Due to uncertainties in the objective function of the optimization problem an analytical treatment based on the derivative is not possible. Recently direct search methods, including Evolutionary Algorithms (EA), proved to be successful for optimization in the presence of noise. A survey on the suitability of EAs for different classes of noisy optimization problems is presented in [13].

Evolution strategies (ES) which build a subclass of EAs that without limitation mainly concentrates on real-valued optimization. The progress of evolution strategies only relies on the function values of the candidate solutions, i.e. on their fitness. An objective function that is subject to noise biases the selection process of the ES which can cause a stagnation of the search process. In order to prevent stagnations the vulnerability of the ES to noise has to be reduced. One can basically apply two methods for reduction of the noise influence on the strategy's performance. On the one hand, the actual fitness of a candidate solution can be replaced by its expected fitness. The expected fitness is computed by multiple evaluations of the same candidate solution and subsequent averaging of the observed function value samples. Increasing the sample size reduces the variance of the estimated fitness provided that the moments of the noise distribution exist. Since fitness evaluations can be expensive, the sample size should be defined as small as possible while simultaneously acquiring a satisfactory performance. Another technique to impair the impact of noise the ES is to increase the population size. In large populations the influence of noise on an individual is likely to be compensated by that of another individual. Recombination also gains a significant performance improvement in noisy fitness environments, see [2]. However, enlarging the population size also implicates an increase of the required number of fitness evaluations.

In order to avoid an excess of function evaluations, the question arises at which point to take the countermeasures, i.e. to increase the population size or to use the averaged function values. Since noise is reflected in changes of a candidate solution's measured fitness between two consecutive evaluations, resampling of fitness values is also suitable for noise detection. But it is not a reliable indicator for progress stagnation. The approach introduced in [12] bases the decision whether to treat the uncertainties or not on an uncertainty level. This level is determined by the number of rank changes within the offspring individuals after applying small perturbations. Although this approach aims at handling so-called actuator noise, it is applicable to fitness noise by replacing the perturbations with an additional fitness evaluation. The work of [9] suggests a population control rule which is based on the residual error. Since the dynamics of an ES in a noisy environment will usually approach a steady state in the vicinity of the optimizer. At that point, fluctuations of the parental fitness values around their mean value can be observed. The population size is then increased if the fitness dynamics on average deny further progress.

This paper investigates the applicability of three different approaches to noisy fitness environments. Two detection techniques are based on linear regression analysis. The first approach estimates the slope of the linear regression line. The direction of the trend can be determined to a certain significance level by computation of a confidence interval on the slope estimator. The second approach uses the Mann-Kendall hypothesis test to infer whether the trend of the strategy's dynamics exhibits increasing or decreasing tendencies. Eventually, the third technique relies on tools known from Time Series Analysis context. In particular, the application of trend decomposition by moving average filters together with the use of specific hypothesis tests on the resulting residual series is regarded.
All approaches are separately integrated into the covariance matrix self-adaptation evolution strategy (CMSA-ES). In particular, the application of trend estimation together with the use of specific hypothesis tests is regarded. The approaches allow for the detection of stagnations within the strategy's progress. Thus the strategy is able to take remedial actions by increasing the population size. The performance of the different strategies is tested on the noisy sphere model as well as the more general noisy ellipsoid model considering three different noise variants. The achieved behavior agrees with existing theoretical predictions [1, 5, 9].

The standard $(\mu/\mu_I, \lambda)$-CMSA-ES is recapped in Sec. 2. Afterwards Sec. 3 introduces the optimization problem and the considered noise models. On this basis the proposed noise handling strategy is presented in Sec. 4. The paper concludes with a discussion of the results.

2 The $(\mu/\mu_I, \lambda)$-CMSA-ES

The methods developed in this paper are integrated into the CMSA-ES. For that reason the standard $(\mu/\mu_I, \lambda)$-CMSA-ES is reviewed at this point. Its pseudo code is displayed in Fig. 1. In each iteration the $(\mu/\mu_I, \lambda)$-CMSA-ES generates λ offspring candidate solutions with its individual mutation strength σ_l. The mutation strength σ can be interpreted as an individual scaling factor. Within the algorithm it is controlled by use of the parameter τ_σ The standard mutation strength learning parameter is chosen $\tau_\sigma = \frac{1}{\sqrt{2N}}$. The mutation vector z_l of each offspring depends on the covariance matrix C which corresponds to the distribution

Initialize	
$g \leftarrow 0$; $\langle\sigma\rangle \leftarrow \sigma^{(init)}$; $\langle \boldsymbol{y}\rangle \leftarrow \boldsymbol{y}^{(init)}$	1
$\mu \leftarrow \mu^{(init)}$; $\lambda \leftarrow \lceil \mu/\nu \rceil$; $C \leftarrow \boldsymbol{I}$	2
Repeat	3
For $l \leftarrow 1$ To λ	4
$\sigma_l \leftarrow \langle\sigma\rangle \mathrm{e}^{\tau_\sigma \mathcal{N}(0,1)}$	5
$\boldsymbol{s}_l \leftarrow \sqrt{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$	6
$\boldsymbol{z}_l \leftarrow \sigma_l \boldsymbol{s}_l$	7
$\boldsymbol{y}_l \leftarrow \langle \boldsymbol{y}\rangle + \boldsymbol{z}_l$	8
$f_l \leftarrow f(\boldsymbol{y}_l)$	9
End For	10
$g \leftarrow g + 1$	11
$\langle \boldsymbol{z}\rangle \leftarrow \sum_{m=1}^{\mu} \boldsymbol{z}_{m;\lambda}$	12
$\langle\sigma\rangle \leftarrow \sum_{m=1}^{\mu} \sigma_{m;\lambda}$	13
$\langle \boldsymbol{y}\rangle \leftarrow \langle \boldsymbol{y}\rangle + \langle \boldsymbol{z}\rangle$	14
$C \leftarrow \left(1 - \frac{1}{\tau_c}\right) C + \frac{1}{\tau_c} \langle \boldsymbol{s}\boldsymbol{s}^\top \rangle$	15
Until $TerminationCriterion$	16

Figure 1: Pseudo code of the $(\mu/\mu_I, \lambda)$-CMSA evolution strategy.

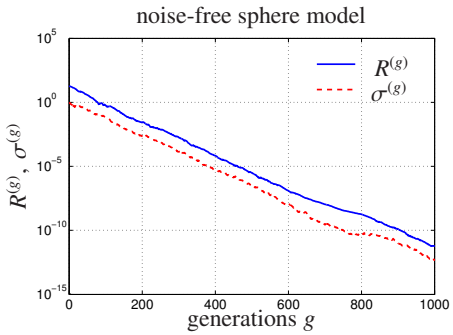

Figure 2: Distance $R^{(g)}$ to the optimizer as well as mutation strength $\sigma^{(g)}$ dynamics of the $(3/3_I, 10)$-CMSA-ES using standard parameter settings. The dynamics are displayed for $N = 40$ on the noise-free Sphere model ($a_i = 1$). The algorithm terminates after 10^3 generations.

of previously generated successful candidate solutions. The update rule can be found in line 15. A standard value for the learning parameter is $\tau_c = 1 + \frac{N(N+1)}{2\mu}$. After construction of the offspring, their objective function (fitness) values are evaluated, see lines 4 to 10.

Having completed the variation step, the algorithm selects those μ of the λ offspring which turn out to have the best fitness values $f_{m;\lambda}$, $m = 1, \ldots, \lambda$. Notice, $m;\lambda$ denotes the mth best out of λ possible values. Accordingly, the notation $\langle . \rangle$ refers to the construction of the centroid of the respective values corresponding to the μ best offspring candidate solutions. For example, the centroid of the mutation strengths is $\langle \sigma \rangle = \frac{1}{\mu} \sum_{m=1}^{\mu} \sigma_{m;\lambda}$.

The algorithm terminates after a predefined termination criterion is met, e.g. maximal number of function evaluations or maximal number of iterations.

3 The noisy fitness environment

The paper focuses on the fitness environment referred to as the ellipsoid model. The according objective function reads

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} a_i x_i^2 \tag{1}$$

with ellipsoid coefficients a_i and search space parameter $\boldsymbol{x} \in \mathbb{R}^N$. A special case of this fitness environment is the frequently studied sphere model $a_i = 1$.

Fig. 2 displays the typical behavior of the $(3/3, 10))$-CMSA-ES with standard learning parameters τ_σ and τ_c on the noise-free sphere model ($a_i = 1, N = 40$). The parental centroid's distance to the optimizer as well as the corresponding mutation strength σ are plotted against the number of generations. The optimization is started at $\mathbf{y}^{(init)} = \mathbf{3}$ with initial stepsize $\sigma = 1$. In the noise-free environment, the CMSA is able to approach the optimizer arbitrarily close ($g \to \infty$) and continuously reduces its mutation strength σ.

Regarding a noisy fitness environment the fitness evaluations are subject to noise. The selection process is then based on the noisy fitness values and might accordingly be biased. We will consider three different types of noise models.

In the first two cases the noise is modeled by an additive term

$$\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \sigma_\epsilon \mathcal{N}(0, 1). \tag{2}$$

That is, a normally distributed noise term with variance σ_ϵ^2 is added to the original fitness of each offspring. The respective standard deviation σ_ϵ is referred to as noise strength.

The models exclude correlations between successive evaluations of Eq. (2). The third noise model is denoted as actuator noise. The actuator noise model reflects internal uncertainties in the variables that are subject to optimization, e.g. actuator imprecision.

The following subsections explain the different noise models and recap existing theoretical results. The effects of different noise models on the dynamics of the same CMSA-ES are illustrated. All results have been obtained by use of the same initial values and strategy parameters as in Fig. 2.

3.1 Constant noise variance

First, consider the noise model of constant noise variance, i.e. the noise strength σ_ϵ is independent from the current location in the search space. In this situation evolution strategies exhibit a common behavior. The noise influence can be neglected far away from the optimizer, i.e. if $f(\boldsymbol{x}) \gg \sigma_\epsilon \mathcal{N}(0,1)$). But as the strategy approaches the optimizer the noise gets more pronounced and the ES finally exhibits some kind of steady state behavior. This steady state on average resides in a certain residual distance from the optimal solution (see also Fig. 3 below).

The expected residual distance on the noisy sphere function is derived in [2, 1]. Its formulation for sufficiently small mutation strength values reads

$$R_\infty = \sqrt{\frac{N\sigma_\epsilon}{4\mu c_{\mu/\mu,\lambda}}}. \tag{3}$$

In this context, the term $c_{\mu/\mu,\lambda}$ refers to the *progress coefficient* of the $(\mu/\mu_I, \lambda)$-ES, [4],

$$c_{\mu/\mu,\lambda} = \tfrac{\lambda-\mu}{2\pi}\binom{\lambda}{\mu}\int_{-\infty}^{\infty} \mathrm{e}^{-t^2}\,(1-\Phi(t))^{\lambda-\mu-1}\,\Phi(t)^{\mu-1}\mathrm{d}t. \tag{4}$$

with $\Phi(t)$ denoting the cumulative distribution function of a standard normal variable.

Considering non-spherical ellipsoid models the derivation of the residual distance becomes more complicated. However, a lower bound for the expected steady state function value on the noisy ellipsoid model is obtained in [5] as

$$\mathrm{E}\left[f_\infty\right] \geq \frac{N\sigma_\epsilon}{4\mu c_{\mu/\mu,\lambda}}. \tag{5}$$

This rather simple result can be explained by the equipartition effect which decomposes the arbitrary directed ellipsoid into its essential fitness components. It was experimentally verified that the equal sign in (5) predicts the steady state fitness well as long as the mutation strength σ of the ES is adapted properly.

The residual distance formula (3) can be extended to non-sperical quadratic models. A residual value R_a for the weighted sum of the parental centroid's components can be defined

$$R_a \coloneqq \sqrt{\sum_{j=1}^{N} a_j^2 y_j^2}. \tag{6}$$

The steady state value R_a^∞ can be determined by means of the noisy quadratic progress rate of the ellipsoid model derived in [17].

$$\varphi_i^{II}(\sigma^{(g)}, \mathbf{y}^{(g)}) = \frac{2\sigma^{(g)} c_{\mu/\mu,\lambda} a_i y_i^{(g)^2}}{\sqrt{(1+\varkappa^2)\sum_{j=1}^{N} a_j^2 y_j^{(g)^2}}} - \frac{\sigma^{(g)^2}}{\mu}, \tag{7}$$

with noise-to-signal ratio $\varkappa$

$$\varkappa^2 = \frac{\sigma_\epsilon^2}{4\sigma^2 \sum_{j=1}^{N} a_j^2 y_j^2}. \tag{8}$$

After the ES has reached its noisy residual steady state distance R_a^∞ to the optimizer, the further progress is zero in expectation. Making use of the noisy quadratic progress rate (7) yields the condition

$$\frac{\sigma^2}{\mu} \overset{!}{=} \frac{2\sigma c_{\mu/\mu,\lambda} a_i y_i^2}{\sqrt{(1+\varkappa^2)\sum_{j=1}^{N} a_j^2 y_j^2}}, \tag{9}$$

The square root in the denominator of Eq. (9) can be transformed into

$$\sqrt{(1+\varkappa^2)\sum_{j=1}^{N} a_j^2 y_j^2} = \frac{\sqrt{4\sigma^2 \sum_{j=1}^{N} a_j^2 y_j^2 + \sigma_\epsilon^2}}{2\sigma}. \tag{10}$$

Thus another representation of (9) is

$$\frac{\sigma^2}{\mu} \overset{!}{=} \frac{4\sigma^2 c_{\mu/\mu,\lambda} a_i y_i^2}{\sqrt{4\sigma^2 \sum_{j=1}^{N} a_j^2 y_j^2 + \sigma_\epsilon^2}}. \tag{11}$$

Multiplying both sides by a_i and summation over all i yields

$$\begin{aligned} \frac{\Sigma a}{4\mu c_{\mu/\mu,\lambda}} &\overset{!}{=} \frac{\sum_{j=1}^{N} a_j^2 y_j^2}{\sqrt{4\sigma^2 \sum_{j=1}^{N} a_j^2 y_j^2 + \sigma_\epsilon^2}} \\ &= \frac{\sum_{j=1}^{N} a_j^2 y_j^2}{\sqrt{\sum_{j=1}^{N} a_j^2 y_j^2}\sqrt{4\sigma^2 + \frac{\sigma_\epsilon^2}{\sum_{j=1}^{N} a_j^2 y_j^2}}} \end{aligned} \tag{12}$$

Notice, that the abbreviation $\Sigma a \coloneqq \sum_{i=1}^{N} a_i$ is used in Eq. (12).

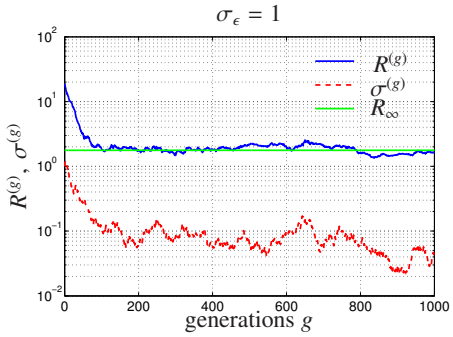

Figure 3: Distance $R^{(g)}$ to the optimizer as well as mutation strength $\sigma^{(g)}$ dynamics of the $(3/3_I, 10)$-CMSA-ES on the noisy sphere model. The dynamics are subject to fitness noise of constant variance. All other settings match those of Fig. 2.

Finally, the residual quantity R_a^∞ can be derived in the limit of small mutation strengths. Assuming small values of σ, the term $4\sigma^2$ in the denominator of (12) can be neglected and one obtains

$$R_a^\infty = \sqrt{\frac{\sigma_\epsilon \Sigma a}{4\mu c_{\mu/\mu,\lambda}}}. \tag{13}$$

This formulation provides a generalization of (3) which can be realized by inserting $a_i = 1$. Equation (13) clearly indicates that increasing the population size is capable of reducing the residual distance to the optimizer (given that the truncation ratio $\vartheta = \mu/\lambda$ is held constant).

The presence of additive fitness noise of constant variance causes the CMSA-ES to approach its residual distance R_a^∞ instead of converging towards the optimizer. This typical behavior is illustrated in Fig. 3 on the noisy sphere model ($a_i = 1$). There, the steady state R_a^∞ is displayed using a solid green line. Being initialized with $\sigma^{(0)} = 1$ at $\boldsymbol{y} = \mathbf{1}$ in dimension $N = 40$ the $(3/3, 10)$-CMSA-ES uses the standard parameter configuration. The solid blue line represents the distance to the optimizer $R^{(g)}$ while the dashed red line illustrates the corresponding mutation strength dynamics. The stagnation of the strategy around its residual steady state R_a^∞ substantiates the idea to enable further progress in direction of the optimizer by increasing the population size. In order to keep the amount of function evaluations as low as possible the right moment for the population augmentation has to be determined. This will be accomplished by detection of stagnations within the strategy's fitness dynamics.

3.2 Constant normalized noise variance

The second noise representation is the model of constant normalized noise variance. On the ellipsoid model the normalized noise strength is defined as [17],

$$\sigma_\epsilon^* \coloneqq \frac{\sigma_\epsilon \Sigma a}{2\sum_{j=1}^N a_j^2 y_j^2}. \tag{14}$$

For $a_i = 1$ this representation includes the well known noise strength normalization on the sphere model [2]

$$\sigma_\epsilon^* \coloneqq \sigma_\epsilon \frac{N}{2R^2}, \tag{15}$$

where $R \coloneqq \|x\|$ is the individual's distance to the optimizer. By definition the noise strength σ_ϵ gets reduced with decreasing distance to the optimizer. That is, in this model the optimizer is noise-free and σ_ϵ increases quadratically with the distance from the optimizer. Distant candidate solutions are subject to huge noise perturbations which can lead the strategy to diverge from the optimizer. Due to its representation (15) this type of fitness noise is also referred to as fitness proportional noise.

According to theoretical investigations on the ellipsoid model [6, 7], positive progress in direction of the optimizer is related to the condition

$$\sigma^{*2} + \sigma_\epsilon^{*2} < 4\mu^2 c_{\mu/\mu,\lambda}^2. \tag{16}$$

The term σ^* denotes the normalized mutation strength which is defined on the ellipsoid model as

$$\sigma^* = \sigma \frac{\Sigma a}{\sqrt{\sum_{j=1}^N a_j^2 y_j^2}}. \tag{17}$$

Notice, that this definition also holds for the sphere model $a_i = 1$. Regarding the limit of small mutation strength σ the condition for positive progress becomes

$$\mu > \frac{\sigma_\epsilon^*}{2c_{\mu/\mu,\lambda}}. \tag{18}$$

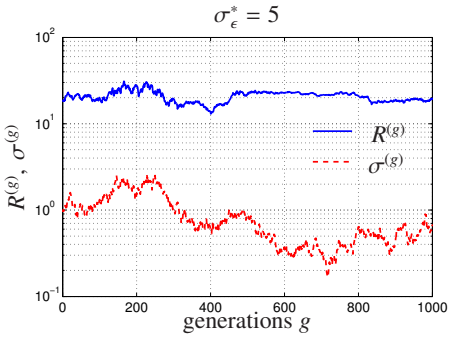

Figure 4: Distance $R^{(g)}$ to the optimizer as well as mutation strength $\sigma^{(g)}$ dynamics of the $(3/3_I, 10)$-CMSA-ES on the noisy sphere model. The noise is modeled according to Eq. (17) as fitness proportional noise. All other settings match those of Fig. 2.

From Eq. (18) it can be deduced that positive progress is connected to the population size. If the strategy diverges, enlarging the population size sufficiently results in returning to positive progress and in the long run in convergence to the noise-free optimizer.

Fig. 4 illustrates the influence of fitness proportional noise on the $(3/3, 10)$-CMSA-ES dynamics. Taking a look at the graphs one observes that the huge normalized noise strength of $\sigma_\epsilon^* = 5$ is preventing the strategy from converging. In this particular case the CMSA-ES rather maintains its initial distance to the optimizer. This does not reflect the general behavior since the impact of strong fitness proportional noise typically results in divergence from the optimizer, cf. the positive progress criterion (18). Conclusively, a procedure which increases the population size in the event of divergence or stagnation would also be beneficial for noise with constant normalized variance.

3.3 Actuator noise

Taking a look at the third noise model it does not directly bias the fitness evaluations. The perturbations affect the search space variables which "manipulate" the fitness values internally. The actuator noise model assigns a noise term to each component of the search space parameter $\boldsymbol{x} \in \mathbb{R}^N$

$$\tilde{f} = f(\boldsymbol{x} + \boldsymbol{\delta}). \tag{19}$$

The components of the random vector $(\boldsymbol{\delta})_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $\forall i = 1, \ldots, N$, are normally distributed with noise variance σ_ϵ^2.

Like in the constant noise variance model, in this scenario the ES is not able to reach the optimizer and will in the long run approach a steady state distance R_∞. Although the asymptotically exact formula has been derived in [8] for the sphere model, for our investigations it is sufficient to provide its approximation for $\sigma \ll \sigma_\epsilon$

$$R_\infty \geq \frac{N\sigma_\epsilon}{\sqrt{8}\mu c_{\mu/\mu,\lambda}} \sqrt{1 + \sqrt{1 + \frac{8\mu^2 c_{\mu/\mu,\lambda}^2}{N}}}. \tag{20}$$

As in the previous cases, this formula shows that larger populations sizes allow for the realization of residual steady state distances closer to the optimizer. A generalization of Eq. (20)

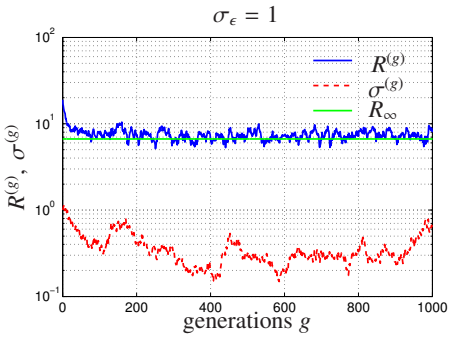

Figure 5: Distance $R^{(g)}$ to the optimizer and mutation strength $\sigma^{(g)}$ dynamics. The $(3/3_I, 10)$-CMSA-ES on the sphere model is illustrated considering the impact of actuator noise of variance $\sigma_\epsilon = 1$. All settings match those of Fig. 2.

for the ellipsoid model with coefficients $a_i \neq 1 \; \forall i$ can be obtained as

$$R_a^\infty \geq \frac{\Sigma a \sigma_\epsilon}{\sqrt{8}\mu c_{\mu/\mu,\lambda}} \sqrt{1 + \sqrt{1 + \frac{8\mu^2 c_{\mu/\mu,\lambda}^2}{\Sigma a}}}. \tag{21}$$

Figure 5 shows the actuator noise influence on the CMSA-ES dynamics. As predicted by Eq. (21) the strategy approaches its steady state distance R_a^∞ which is displayed by the solid green line. Except for the larger residual distance to the optimizer the typical behavior of the CMSA-ES under actuator noise closely resembles the additive noise case of constant variance, cf. Fig. 3. Regarding the dependency of Eq. (21) on the population size one is also interested in a mechanism which identifies stagnations within the strategy's fitness dynamics. Hence, the residual distance can be successively decreased and long stagnations can be prevented.

4 Noise detection

This section aims at the presentation of mechanisms which are able to cope with noise. The underlying idea is to identify stagnations in the fitness value dynamics of the CMSA-ES which are caused by the three noise models mentioned in Sec. 3. The detection mechanism is based on the fitness of the parental centroid resulting in one additional function evaluation per iteration. Having recognized a significant stagnation of the parental centroids fitness we are able to apply counteractions, i.e. increase the population size.

Interpreting the parental fitness dynamics as a time series its trend can be analyzed. The first detection method uses a linear regression model of the fitness sequence. A confidence interval of the model's slope can be used to test if the slope of the estimated linear regression line is different from zero. Negative trend indicates further progress in direction of the optimizer while non-negative trend points towards stagnation or even towards divergence away from the optimizer.

Another mechanism for trend evaluation is the Mann-Kendall hypothesis test, [16, 14]. It can be used in place of a parametric linear regression analysis. The Mann-Kendall test statistically assess if there is a monotonic downward trend of the fitness values. A monotonic downward trend indicates that the fitness consistently decreases through time, but the trend may or may not be linear.

The third approach to identify fitness stagnations is realized by methods from the context of time series analysis, [10]. Interpreting the parental fitness dynamics as a time series, its trend can be estimated by a moving average filter or exponential smoothing, respectively. The elimination of the trend components from the fitness sequence yields a residual series which is tested for significant autocorrelations. The absence of autocorrelations within the residual series is then interpreted as indicator for stagnation in a certain steady state distance from the optimizer. We refer to this approach as the residual decision method.

The considered techniques are recapped and discussed in the next subsections. Finally, they are applied to a sequence of the typical noisy fitness dynamics resulting from the CMSA-ES, see Fig. 1. Considering the different noise models, all three approaches are able to detect stagnations within the noisy fitness dynamics.

4.1 Linear regression

Linear regression is the least squares estimator of a regression model with a single explanatory variable. That is, a straight line is fitted through a set of n data points $(x_1, y_1), \ldots, (x_n, y_n)$ in such a way that the sum of squared residuals of the model is minimal. The slope of the fitted line is equal to the correlation between y and x corrected by the ratio of standard deviations of these variables. The intercept ensures that the fitted line passes through the center of mass (x, y) of the data points. By derivation of a confidence interval for the estimated slope one is able to find an indicator for a negative linear trend.

Suppose there are n data points $\{(x_i, y_i),\ i = 1, \ldots, n\}$. The function that describes x and y is represented by

$$y_i = ax_i + b + \epsilon_i. \tag{22}$$

Being interested in the straight line

$$y = ax + b \tag{23}$$

that provides a "best" fit for the data points in terms of the least-squares approach, one searches a line that minimizes the sum of squared residuals of the linear regression model. Accordingly, to obtain the y-intercept b and the slope a, the following minimization problem has to be solved,

$$\min_{a,b} \sum_{i=1}^{n} (y_i - b - ax_i)^2. \tag{24}$$

Using calculus, cf. [15], the least squares estimators for a and b can be derived as

$$\hat{a} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{25}$$

and

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \tag{26}$$

where $\bar{x}$ and $\bar{y}$ represent the sample mean values of the observations. The formulas (25) and (26) allow for the calculation of the coefficients a and b of the regression line for the given data set. In order to make an assertion about the preciseness of the estimation the corresponding confidence interval is computed. The confidence interval determines a plausible set of values for the estimates given that the experiment is repeated a very large number of times.

Assuming that the number of observations n is sufficiently large, the central limit theorem guarantees that the estimator of the slope is approximately normally distributed with mean a. Under the normality assumption the sum of squared residuals $\sum_{i=1}^{n}(y_i - b - ax_i)^2$

is distributed proportionally to χ^2_{n-2} with $n-2$ degrees of freedom, and independent from $\hat{a}$ [15]. This allows to construct a t-statistic

$$t = \frac{\hat{a} - a}{s_{\hat{a}}} \sim t_{n-2}, \tag{27}$$

where

$$s_{\hat{a}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - b - ax_i)^2}{(n-2)\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{28}$$

represents the standard error of the estimator $\hat{a}$. The t-statistic has a Student's t-distribution with $n-2$ degrees of freedom. On this basis a confidence interval for a can be constructed

$$a \in [\hat{a} - s_{\hat{a}} t^{\alpha}_{n-2},\ \hat{a} + s_{\hat{a}} t^{\alpha}_{n-2}] \tag{29}$$

with confidence level $(1-\alpha)$, where t^{α}_{n-2} is the $(1-\alpha/2)$-th quantile of the t_{n-2} distribution. The confidence interval for a gives a general idea in which range the slope of the regression line is most likely located. In situations where the upper boundary of the confidence interval is smaller or equal to zero a negative linear trend within the set of observations can be expected. Conclusively, this leads to the condition

$$\hat{a} + s_{\hat{a}} t^{\alpha}_{n-2} < 0. \tag{30}$$

On the other hand the violation of this condition does not allow for a proposition of the trend direction. The absence of a significant negative trend within the ES dynamics can thus be interpreted as an indicator for the presence of progress stagnations.

Linear regression analysis is applied to a series of noisy fitness dynamics of a $(3/3_I, 10)$-CMSA-ES in Fig. 6. The illustration displays the negative test decisions for a negative linear trend on intervals of $3N$ fitness observations for $\alpha = 0.05$. A non-negative trend is used as an indicator for noise related progress stagnations. The linear regression approach is integrated into the CMSA-ES in Sec. 5.

4.2 The Mann-Kendall test

While the regression analysis requires that the residuals from the fitted regression line are normally distributed this assumption is not required by the Mann-Kendall test. The Mann-Kendall test (MK test) is a non-parametric test, i.e. it does neither depend upon the magnitude of data nor on assumptions of the distribution[1]. It assesses whether a time-ordered data set exhibits an increasing or decreasing trend, within a predetermined level of significance. The MK test is based on the assumption that the absence of a trend indicates that the measurements obtained over time are independent and identically distributed. The assumption of independence means that the observations are not autocorrelated.

According to [11] the first step to compute the MK test statistic is to list the data in their order of appearance

$$x_1, x_2, \ldots, x_n, \tag{31}$$

and to determine the sign of all $n(n-1)/2$ possible differences

$$x_j - x_k, \quad j > k. \tag{32}$$

[1] Thus, it could be also applied to heavy tail noise.

The sign of $x_j - x_k$ indicates that the observation x_j at time j is greater or smaller than the observation x_k at time k. The next step is concerned with the computation of the number of positive differences minus the number of negative differences

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sign}(x_j - x_k). \tag{33}$$

A positive value S corresponds to the situation that observations obtained later in time tend to be larger than observations made earlier. On the other hand a negative S value indicates that the observations decrease with time. For sample size $n > 10$ a trend in the data can be determined by computation of the variance of S

$$\operatorname{Var}[S] = \frac{n(n-1)(2n+5)}{18}. \tag{34}$$

Afterwards, the (normally distributed) Mann-Kendall test statistic is derived

$$Z_{mk} = \begin{cases} \frac{S-1}{\sqrt{Var[S]}}, & \text{if } S > 0, \\ 0, & \text{if } S = 0, \\ \frac{S+1}{\sqrt{Var[S]}}, & \text{if } S > 0. \end{cases} \tag{35}$$

A positive or negative value of Z_{mk} indicates that the data tend to increase or decrease, respectively. In order to test the null hypothesis

$$H_0: \quad \text{No monotonic trend} \tag{36}$$

against the alternative hypothesis

$$H_a: \quad \text{Downward monotonic trend} \tag{37}$$

at the significance level α, H_0 is rejected and H_a is accepted if

$$Z_{mk} \leq Z_{1-\alpha}. \tag{38}$$

That is, the value of the Mann-Kendall test statistic is lower or equal than $Z_{1-\alpha}$, the $(1-\alpha)$ percentile of the standard normal distribution.

An application of the MK test to a series of noisy fitness dynamics of a $(3/3_I, 10)$-CMSA-ES is illustrated in Fig. 6. Again the test decisions which reject a significant negative trend are displayed and used to indicate possible progress stagnations with $\alpha = 0.05$. The Mann-Kendall test is incorporated into the CMSA-ES in Sec. 5.

4.3 Residual decision

The residual decision approach is a two-step procedure which decomposes the time series of observations into a trend and a residual component, respectively. The residual series is then analyzed for significant auto-correlations in the second step. The absence of auto-correlations between the residual values is regarded as an indicator for stagnations.

4.3.1 Time series decomposition

Given a sequence of observations it can be regarded as a time series in the form of the decomposition model

$$x_t = m_t + r_t, \quad t = 1, \ldots, n, \tag{39}$$

with trend component m_t and residual component r_t. Notice, that seasonal dependencies are ignored at this point as their appearance is not expected in the present applications.

The corresponding trend m_t can be estimated by application of a finite two-sided moving average filter. For a nonnegative integer q the moving average provides the trend estimates

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} x_{t-j}, \qquad q+1 \leq t \leq n-q. \tag{40}$$

The filter removes rapid fluctuations from the original time series x_t and leaves the slowly varying estimated trend term $\hat{m}_t$. This estimation assumes that the trend m_t of the observations x_t is approximately linear within the interval $[t-q, t+q]$. Since x_t is not observed for time steps $t \leq 0$ or $t > n$, the estimated trend sequence is $2q$ entries shorter than the original observation series.

There are multiple other filters that can be used for smoothing the time series. Another way is the application of exponential smoothing. For a predefined $\xi \in [0, 1]$, it generates the one-sided moving averages by the recursions

$$\begin{aligned} \hat{m}_t &= \xi x_t + (1-\xi)\hat{m}_{t-1}, \quad t = 2, \ldots, n, \\ \hat{m}_1 &= f_1. \end{aligned} \tag{41}$$

Except for the last one referring to f_1, the weights of this recursion decrease exponentially.

Having estimates the trend of the fitness dynamics, the corresponding residual series $\hat{r}_t,\ q+1 \leq t \leq n-q$ can be computed

$$\hat{r}_t = f_t - \hat{m}_t, \qquad q+1 \leq t \leq n-q. \tag{42}$$

The next step is concerned with testing this estimated residual sequence $\hat{r}_t$ for autocorrelation.

4.3.2 Hypothesis test

The algorithm tries to identify significant noise influence by analysis of the residual time series. For that reason the residuals are tested for autocorrelation. If there exists a relation between the entries of the sequence which is more than random, the measured autocorrelation typically has a value that is significantly different from zero. The observations within the sequence are then referred to as autocorrelated. This section recaps the Ljung-Box Q-test for checking the hypothesis that the residuals r_t are realizations of independent and identically distributed random variables and thus show no significant autocorrelation. If the hypothesis is true, measuring the mean and variance of the noise sequence suffices to describe the distribution of the residuals. Thus it is likely that the residual series describes the fluctuations around a constant trend which indicates a stagnation in the strategy's progress.

The considered tests use the sample auto-correlation function for the hypothesis check whether the residual series is auto-correlated up to lag h or not. The sample auto-correlation of a set of realizations is defined as follows

Definition 1. *Let $(x_t)_{t\in I}$ ($I = \{1, \ldots, n\}$) be the observations of a time series. The **sample mean** of (x_t) is given by*

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t. \tag{43}$$

*The **sample autocovariance function** is then defined as, [10],*

$$\hat{\gamma}_X(h) \coloneqq \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|}-\bar{x})(x_t-\bar{x}), \quad -n<h<n, \tag{44}$$

*and the **sample autocorrelation function (ACF)** of (X_t) at lag h becomes*

$$\hat{\rho}(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}, \quad -n<h<n. \tag{45}$$

Based on the sample autocorrelation function of the time series a hypothesis test for identification of autocorrelation within the residual series is introduced.

The Ljung-Box Q-test The Ljung-Box Q-test is a modification of the classical portmanteau test statistic proposed by Box and Pierce [10]. It allows to test for autocorrelation at multiple lags jointly. The null hypothesis that the first h autocorrelations are jointly zero,

$$H_0 : \hat{\rho}(1) = \hat{\rho}(2) = \cdots = \hat{\rho}(h) = 0, \tag{46}$$

is tested against the alternative hypothesis that the observations exhibit serial correlation. The Ljung-Box test statistic is calculated as

$$Q_{LB}(h) = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}(k)^2}{n-k}. \tag{47}$$

where $\hat{\rho}(k)$ is the sample autocorrelation of order k of the residuals. Under the null hypothesis, $Q_{LB}(h)$ follows a χ^2 distribution with h degrees of freedom. That is, considering a significance level of α, the null hypothesis is rejected if

$$Q_{LB} > \chi^2_{1-\alpha}, \tag{48}$$

where $\chi^2_{1-\alpha}$ denotes the α-quantile of the χ^2 distribution with h degrees of freedom. The test performance depends on the choice of h. If n is the number of observations, choosing $h = \ln(n)$ is recommended. The Ljung-Box Q-test can be used to test for autocorrelation in any series with a constant mean. This includes the residual series constructed in section 4.3.1.

It should be noted that the residual decision approach does only indicate the presence of stagnations. That is, the Ljung-Box test will reject the null hypothesis that the residuals are iid regardless whether there exists either a negative trend or a positive one. On the other hand the linear regression trend analysis as well as the Mann-Kendall test explicitly examine the presence of a significant negative trend.

The application of the residual decision method to a series of noisy fitness dynamics of a $(3/3_I, 10)$-CMSA-ES is displayed in Fig. 6. Interpreting the fitness values as a time series in the form of the decomposition model

$$\tilde{f}_t = m_t + r_t, \quad t = 1, \ldots, L \tag{49}$$

the corresponding trend m_t and residual series r_t can be analyzed. Fig. 6 considers the noise-free ellipsoid model $a_i = i$ in dimension $N = 30$ as well as the three noise models discussed in Sec. 3. The trend is derived by use of the exponential smoothing filter with $\xi = 1/\sqrt{2N}$. The hypothesis tests are then applied to the residual series in the following way: Once the CMSA-ES has established a series of $L = 3N$ fitness values and residuals, respectively, the interval of the last L residuals is tested for autocorrelation using the Ljung-Box test (with lag $h = \lfloor \ln(L) \rfloor$ and $\alpha = 0.05$). If the test indicates that the residuals within the interval shows no

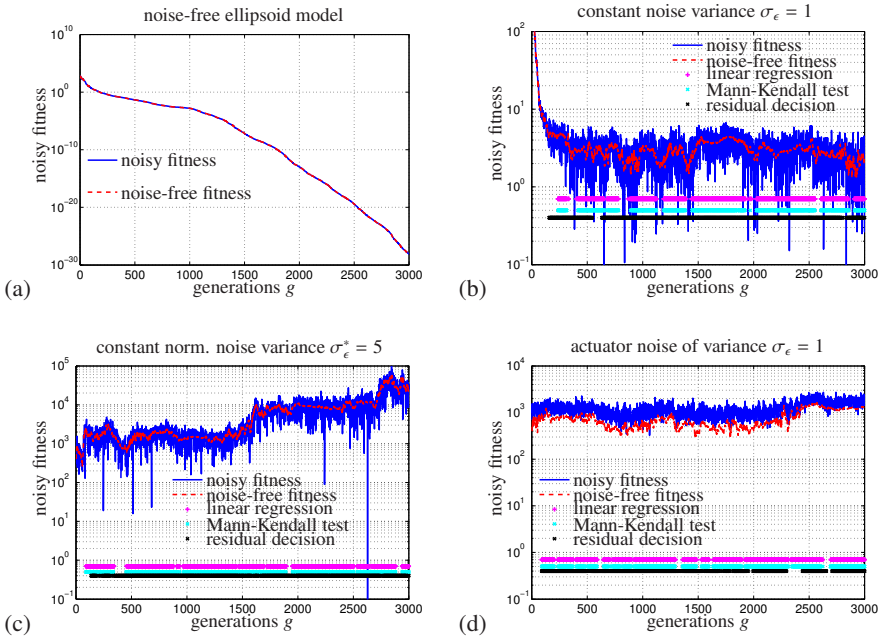

Figure 6: The fitness dynamics of a $(3/3_I, 10)$-CMSA-ES run on the ellipsoid model ($a_i = i$) using standard parameter settings in dimension $N = 30$. The noise-free case as well as the three different noise models are displayed from (a) to (d). The test decisions indicating significant evidence for a present stagnation are illustrates by the data points.

significant autocorrelations, this suggests that the strategy's progress is decreasing resulting in more evenly distributed residuals around the estimated trend of the fitness observations. Such positive test results for stagnation are indicated by the black "x" data points. The linear regression approach and the MK test are realized on the same fitness intervals of length $L = 3N$. Their test decisions indicating stagnations are illustrated by the magenta "$\star$" as well as the light blue "x" data points, respectively.

In the noise-free case (a) none of the three approaches indicates stagnations. This does not come as a surprise since the fitness observations are not biased by noise. Linear regression as well as Mann-Kendall recognize a significant negative trend within the examined fitness intervals. The residual decision method also indicates a significant negative trend which corresponds to the fact that the residuals still show autocorrelations.

Considering constant noise variance σ_ϵ in (b), the residual sequence appears to be nearly stationary after a sufficiently long time. After about 100 generations the progress of the strategy slows down considerably. This behavior corresponds to the strategy approaching its steady state distance to the optimizer. As a result the estimated trend of the fitness observations is increased until its slope fluctuates around zero. Hence, linear regression as well as the MK test fail to detect a significant negative trend. Considering the residual decision technique the estimated trend represents the average fitness more precisely and the residuals appear to be uncorrelated. After the first positive test decision for non-negative trend or zero autocorrelation, respectively, the test decisions are confirmed on almost all following intervals of length $3N$.

Similar results are observable in (c) for constant normalized noise variance σ^*_ϵ. There,

the residual components are rated as uncorrelated after the fitness dynamics of the CMSA-ES do no longer show significant progress. All detection mechanisms observe the stagnation within the fitness dynamics of the evolution strategy. In the beginning the residual decision approach needs a slightly longer phase until it indicates progress stagnation. On the other hand it seems more robust towards short intervals with a significant negative trend. Due to high fluctuations within the fitness dynamics these intervals of downward trend can cause the linear regression analysis as well as the Mann-Kendall test to detect no noise influence. This is reflected by the small gaps, e.g. between generation 400 and 500, in the data point plots of the test decisions.

The actuator noise case is displayed in (d). In that situation the fitness dynamics are subject to perturbations within the search space parameters. Again all three detection mechanisms observe the stagnation within the fitness dynamics of the evolution strategy. Alike the constant noise case the fitness dynamics do approach a residual steady state, but in a greater distance from the optimizer. After the transient phase almost all test decisions indicate the presence of a fitness stagnation caused by actuator noise. The occurrence of single opposite decisions may potentially be caused by large deviations within the noisy fitness dynamics. The illustration in Fig. 6 verifies the use of the three proposed detection mechanisms for identification of stagnations within the fitness dynamics of an evolution strategy.

5 Application to CMSA-ES

This section aims at the construction of algorithms which are able to cope with fitness noise. Therefore, the methods specified in Sec. 4 are incorporated into the basic $(\mu/\mu_I, \lambda)$-CMSA-ES, see Fig. 1. The adjusted algorithm tries to identify significant noise influence that causes a stagnation of the strategy progress. For that reason the fitness dynamics of the evolution strategy are examined. Having encountered a significant stagnation the algorithm will increase the populations sizes μ and λ while keeping the truncation ratio $\vartheta = \lambda/\mu$ constant. The population control covariance matrix self-adaptation evolution strategy, (μ, λ)-pcCMSA-ES, is illustrated in Fig. 7.

Until the algorithm has generated a list $\mathcal{F}$ of L parental function values the CMSA-ES remains unchanged. Subsequently, the pcCMSA-ES examines the list $\mathcal{F}$ using the methods explained in Sec. 4, i.e. either linear regression, the Mann-Kendall test, or residual decision. These procedures are represented by the single program `detection`$(\mathcal{F}_{int}, \alpha)$, line 20. Analyzing the fitness interval $\mathcal{F}_{int}$ in order to identify stagnations, it returns the decision variable $td = 1$ as long as a negative trend among the fitness samples is estimated. Otherwise it returns the test decision $td = 0$. The parameter α refers to the significance level of the underlying hypothesis tests.

Provided that a negative trend is detected the algorithm acts like the original CMSA-ES. It controls the covariance matrix C of the offspring distribution as well as the parental mutation strength in the usual way. However, if the initial population size has been increased within an earlier iteration, then it is reduced again in line 25. This ensures that the algorithm is able to readjust the population size once the strategy has left noisy search space regions.

After having found stagnations within the fitness sequence which are associated with non-negative trend ($td = 0$), the algorithm should be able to apply corrective actions in order to ensure a continuing approach to the optimizer. Accordingly, the population sizes are increased by the factor c_μ, line 22, keeping the truncation ratio $\vartheta = \frac{\mu}{\lambda}$ constant. Having adjusted the population size, the detection procedure is interrupted for L generations (line 23). Otherwise the basis of the next test decision would be biased by the old fitness values generated by use of smaller populations. Additionally the covariance matrix adaptation is turned off in line 24, once the algorithm has encountered significant noise impact. For this

Initialization
$g \leftarrow 0$; $q \leftarrow 0$; $\langle\sigma\rangle \leftarrow \sigma^{(init)}$; $\langle \mathbf{y}\rangle \leftarrow \boldsymbol{y}^{(init)}$; $\mu \leftarrow \mu^{(init)}$ 1
$\mu_{\min} \leftarrow \mu^{(init)}$; $C \leftarrow \text{eye}(N)$; ϑ; L; $adjC \leftarrow 1$ 2
Repeat 3
$\lambda \leftarrow \lfloor \mu/\vartheta \rfloor$ 4
For $l \leftarrow 1$ To λ 5
$\sigma_l \leftarrow \langle\sigma\rangle \mathrm{e}^{\tau_\sigma \mathcal{N}(0,1)}$ 6
$\boldsymbol{s}_l \leftarrow \sqrt{C}\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ 7
$\boldsymbol{z}_l \leftarrow \sigma_l \boldsymbol{s}_l$ 8
$\boldsymbol{y}_l \leftarrow \langle \boldsymbol{y}\rangle + \boldsymbol{z}_l$ 9
$f_l \leftarrow f(\boldsymbol{y}_l)$ 10
End For 11
$g \leftarrow g + 1$ 12
$\langle \boldsymbol{z}\rangle \leftarrow \sum_{m=1}^{\mu} \boldsymbol{z}_{m;\lambda}$ 13
$\langle\sigma\rangle \leftarrow \sum_{m=1}^{\mu} \sigma_{m;\lambda}$ 14
$\langle \boldsymbol{y}\rangle \leftarrow \langle \boldsymbol{y}\rangle + \langle \boldsymbol{z}\rangle$ 15
Add $f(\langle \boldsymbol{y}\rangle)$ **To** $\boldsymbol{\mathcal{F}}$ 16
If $g > L \quad \wedge \quad wait = 0$ 17
$\boldsymbol{\mathcal{F}}_{int} \leftarrow \boldsymbol{\mathcal{F}}(g - L : g)$ 18
$td \leftarrow \texttt{detection}(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ 19
If $td = 0$ 20
$\mu \leftarrow \mu c_\mu$ 21
$wait \leftarrow L$ 22
$adjC \leftarrow 0$ 23
Else 24
$\mu \leftarrow \max\left(\mu_{\min}, \lfloor \mu/(1 + \frac{1}{c_\mu})\rfloor\right)$ 25
End If 26
Else 27
$wait \leftarrow wait - 1$ 28
End If 29
$C \leftarrow \left(1 - \frac{1}{\tau_c}\right)^{adjC} C + \frac{adjC}{\tau_c}\langle \boldsymbol{s}\boldsymbol{s}^\top\rangle$ 30
Until $TerminationCriterion$ 31

Figure 7: The (μ, λ)-pcCMSA-ES for noisy fitness environments. The methods for the detection of progress stagnations are implemented into the subroutine $\texttt{detection}(\boldsymbol{F}_{int}, \alpha)$ which is specified within the following subsections.

purpose the parameter $adjC$ is set to zero which stops the covariance matrix update in line 30. Proceeding like this prevents further rise of the conditioning number of the covariance matrix without gaining any useful information from the noisy environment. The algorithm repeats until it reaches its termination criterion.

This first variant of the (μ, λ)-pcCMSA-ES, Fig. 7, spares covariance matrix adaptation after significant noise influence has been identified for the first time. Thinking of fitness environments where the strategy has to deal with noisy regions, it might be beneficial to turn the covariance matrix adaptation back on once the ES has overcome the interferences. That is, if a significant negative trend is present again. The parameter $adjC$ should then be reset to one in order to gain additional information about advantageous search directions. This can easily be obtained by inserting

$$adjC \leftarrow 1$$

between the lines 24 and 25 of the pcCMSA-ES algorithm in Fig. 7. Regarding the noisy fitness environments considered here, the adjustment was not able to provide meaningful improvements in terms of the strategy's progress and was therefore omitted in the present work. However, the contribution of the respective adjustment has to be evaluated in additional studies concerning appropriate noisy fitness environments.

Concentrating on the implementation of the subroutine `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ in line 20 of Fig. 7 the three different detection mechanisms are considered. The first algorithm variant detects stagnations within the fitness dynamics based on linear regression trend analysis. Accordingly, the resulting algorithm is referred to as (μ, λ)-pcCMSA$_{lr}$-ES. The approach transfers the method from Sec. 4.1 to the time series of the last L noisy fitness observations which are gathered within $\boldsymbol{\mathcal{F}}_{int}$. This allows to compute the confidence interval for the estimator of the slope of the fitness trend to a predefined significance level α. Given that the upper boundary of the confidence interval is less than zero this indicates a negative trend within the fitness observations with an error probability of at most α. Accordingly, the function `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ returns the boolean value 1. On the other hand, an upper boundary which is greater or equal to zero allows no distinct decision about the trend direction. This uncertainty can be interpreted as an indicator for the presence of progress stagnations caused by noise. As `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ returns the boolean value 0, the strategy is required to increase the population size.

Applying the Mann-Kendall test within `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ (line 20, Fig. 7) for the purpose of trend estimation the respective algorithm is denoted (μ, λ)-pcCMSA$_{mk}$-ES. The underlying test is specified within Sec. 4.2. It is also applied to the time series of the last L noisy fitness observations stored in $\boldsymbol{\mathcal{F}}_{int}$. Alike using linear regression trend analysis, the subroutine `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ using Mann-Kendall returns the boolean values 1 or 0 if a significant negative trend is present, or not, respectively.

Considering the decomposition of the noisy fitness dynamics $\tilde{f}_t = m_t + r_t$ into trend and residual components results in the third method for identification of progress stagnations. According to Sec. 4.3 the program `detection`$(\boldsymbol{\mathcal{F}}_{int}, \alpha)$ relies on the Ljung-Box hypothesis test which is applied to the residual series r_t. The test returns the boolean value 1 if the null hypothesis that the residual series is independent and identically distributed is accepted. This suggests that the residuals are evenly fluctuating around a stable value without significant trend. Thus the test decision on the autocorrelation between the residuals can be used as an indicator for progress stagnation. Applying the residual decision method to the CMSA-ES we referred to the corresponding algorithm as $(\mu/\mu_I, \lambda)$-pcCMSA$_{rd}$-ES.

6 Empirical comparison of the algorithm variants

Regarding the proposed algorithm variants their performance is compared in this section. Particularly the pcCMSA$_{lr}$-ES, the pcCMSA$_{mk}$-ES, as well as pcCMSA$_{rd}$-ES are confronted with minimization on different noisy fitness environments. To this effect we apply the three noise models introduced in Sec. 3 to the ellipsoid models determined by $a_i = 1, \forall i = 1, \ldots, N$ and $a_i = i,\ \forall i = 1, \ldots, N$, respectively.

All three pcCMSA-ES variants are initialized with standard parameter settings and $\sigma^{(init)} = 1$ at $\boldsymbol{y}^{(init)} = \mathbf{1}$ with search space dimension $N = 30$. The initial population sizes are set to $\mu = 3$ and $\lambda = 9$. The corresponding truncation ratio $\vartheta = \frac{\mu}{\lambda}$ is kept constant during the runs. In the case that progress stagnations are observed by the algorithms the population adaptation parameter $c_\mu = 2$ is used to increase the populations. The hypothesis tests use the parameter $\alpha = 0.05$.
After the population size has been increased the algorithms wait for a predefined period of iterations before the next interval of fitness values $\mathcal{F}_{int}$ is tested for stagnation. The length

of the waiting period as well as the length of the fitness interval $\mathcal{F}_{int}$ is controlled by the parameter L. For the comparisons it is set to $L = 3N$ on the sphere model and to $L = 5N$ on the ellipsoid model.

On the sphere model the algorithms are equipped with a budget of maximal $2 \cdot 10^6$ function evaluations or are iterated over a maximum of 5000 generations. For the runs on the ellipsoid model the function evaluation budget is raised to $3 \cdot 10^6$. To guarantee an unbiased comparison the random number generator settings are saved and restored before the execution of each algorithm.

All illustrations compare four characteristic dynamics. In (a) the noise-free parental fitness dynamics are displayed. The ratio between the logarithm of the fitness and the logarithm of the number of aggregated function evaluations is plotted against the logarithm of aggregated function evaluations. This representation of the fitness dynamics is chosen to allow a connection to the convergence rate investigations on noisy optimization problems developed by Astete-Morales *et al.* [3]. Considering additive noise on quadratic functions, the authors claim that the respective quantity does not fall below a value of $-\frac{1}{2}$, i.e.

$$\frac{\log\left(F(\langle \mathbf{y} \rangle)\right)}{\log(\#\text{ function evaluations})} > -\frac{1}{2}.$$

Regarding the sphere model subject to noise of constant variance our experiments suggest that the pcCMSA-ES is able to come below the proposed boundary. This behavior has been observed provided that the budget of function evaluations is large enough for the algorithm to steadily increase the population sizes. This observation is part of further examinations and will be discussed in more detail within future studies. We also use the same representation for the other noise models in order to ensure comparability of the fitness dynamics.

Figure (b) displays the dynamical behavior of the quantity $\sqrt{\sum_{j=1}^{N} a_j^2 y_j^2}$. For the sphere model $a_i = 1$ this value simply represents the distance of the parental population centroid $\langle y \rangle$ to the optimizer. In the case of the ellipsoid model $a_i = i$ it can be interpreted as a weighted distance defined through the ellipsoid coefficients. Additionally, the plots (c) and (d) display the mutation strength dynamics as well as the normalized mutation strength dynamics. The solid blue line always represents the dynamics of the pcCMSA$_{lr}$-ES which uses linear regression analysis for stagnation identification. The dynamics of the pcCMSA$_{mk}$-ES applying the Mann-Kendall test and the residual decision variant referred to as pcCMSA$_{rd}$-ES are illustrated by use of solid red lines and green lines, respectively.

6.1 Constant noise variance

The first noisy fitness environment considered is the one governed by additive noise of constant variance $\sigma_\epsilon = 1$. Fig. 8 illustrates the algorithm dynamics on the noisy sphere model.

We observe that all three algorithm variants qualitatively exhibit the same dynamics on both fitness environments. On the sphere model in Fig. 8 pcCMSA$_{lr}$-ES and pcCMSA$_{mk}$-ES actually act similarly. Considering the ellipsoid model ($a_i = i$) the detection mechanisms within the CMSA-ES appear more diverse.

The noise-free fitness dynamics, see (a), are continuously decreased with growing number of function evaluations. Regarding (b) we deduce that the algorithms indeed identify stagnations and accordingly increases the population size after a respective test decision ($td = 0$). Under the influence of fitness noise with constant variance the adjusted CMSA-ES algorithms are following the prediction of the residual steady state distance R_∞ to the optimizer, see Eq. (20). Considering the ellipsoid model in Fig. 9, the strategy approaches the R_a^∞ in (21), respectively. Both quantities depend on the parental population size μ as well as the truncation ratio $\vartheta = \frac{\mu}{\lambda}$. As the strategies increase μ, while keeping ν constant,

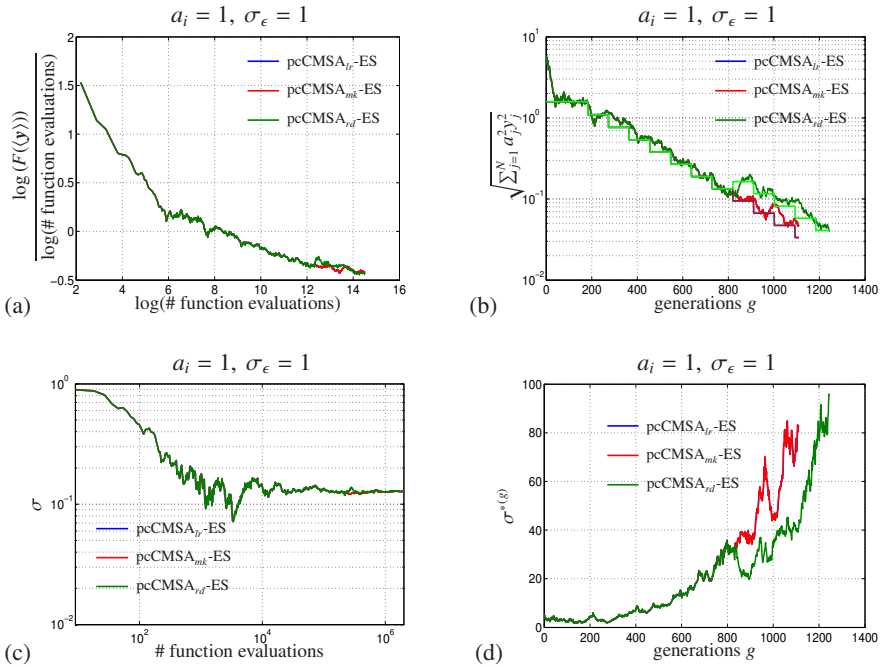

Figure 8: Comparison of the pcCMSA-ES variants, Fig. 7, on the noisy sphere model with noise of constant variance $\sigma_\epsilon = 1$. The normalized mutation strength σ^* is calculated according to Eq. (16).

the residual distance to the optimizer is reduced repeatedly. The predictions of the residual distances for the CMSA-ES variants are illustrated within Fig. 8 and 9, respectively, by use of the equally colored step functions.

Taking a look at the mutation strength dynamics in (c) the σ dynamics approach a steady state. The corresponding normalized mutation strength dynamics are increasing as the strategies approach the optimizer. This compensatory behavior is typical for a self-adaptive ES which has to deal with relatively huge noise perturbations [18].

However, the original CMSA-ES in Fig. 1 would only approach the residual distance governed by the initial population sizes ($\mu = 3, \lambda = 9$). Conclusively, it is not able to generate candidate solutions comparably close to the optimizer. However, due to the significant increase of the population size, the adjusted CMSA-ES variants consume considerably more function evaluations to achieve their improved performances. Figure 9(b) shows that the pcCMSA$_{rd}$-ES increases its populations faster than the other strategies. Thus in terms of iterations it advances faster in direction of the optimizer but on the other hand the huge populations sizes cause the pcCMSA$_{rd}$-ES to terminate after about 2000 generations. This is due to the predefined budget of function evaluations which is exceeded approximately 750 generations earlier than those of the pcCMSA$_{mk}$-ES. In the end, the three pcCMSA-ES variants approach the optimizer at a similar distance.

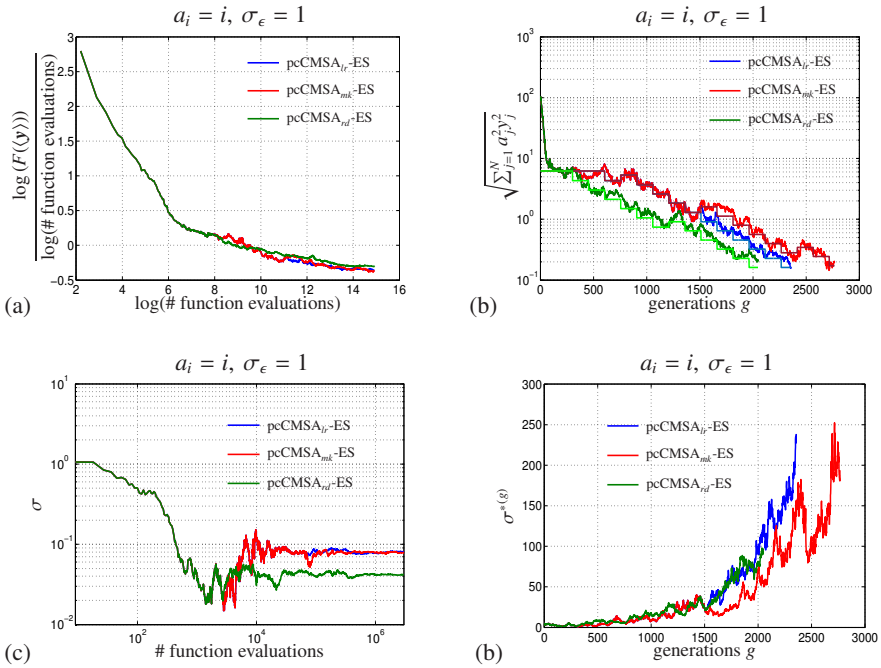

Figure 9: Comparison of the pcCMSA-ES variants on the noisy ellipsoid model with noise of constant variance $\sigma_\epsilon = 1$. The normalized mutation strength σ^* is calculated according to Eq. (16).

6.2 Constant normalized noise variance

Considering additive fitness noise of constant normalized noise variance σ_ϵ^* the dynamics of the pcCMSA-ES variants are compared on the sphere and ellipsoid model in Fig. 10 and Fig. 11, respectively. The results on the sphere model and on the ellipsoid model with $a_i = i$ are illustrated in Fig. 10. Both $(3/3, 10)$-CMSA-ES are initialized with standard parameter settings and $\sigma^{(init)} = 1$ at $\mathbf{y}^{(init)} = \mathbf{3}$. After the first strategy the random number generator settings are saved and restored before the execution of the second algorithm in order to provide comparable runs. Regarding the sphere model the fitness dynamics decrease rapidly once the strategy has reached a parental population size that satisfied condition (18). This is reflected by the $\sqrt{\sum_{j=1}^{N} a_j^2 y_j^2}$ dynamics in (b). The step function plots in (b) also illustrate the dynamics of the offspring population size λ. Again the pcCMSA$_{lr}$-ES and the pcCMSA$_{mk}$-ES result in similar dynamics. One observes that the pcCMSA$_{rd}$-ES realizes slightly larger populations. While it may consume more function evaluations this behavior again leads to a faster approach to the optimizer. The intervals of very steep decent stem from periods where the pcCMSA$_{rd}$-ES adapts larger populations sizes than the other two algorithm variants. Further, the λ-dynamics show that the populations sizes within all pcCMSA-ES variants fluctuate around a rather small value. This illustrates that the strategies are able to adapt the appropriate populations sizes needed to comply with Eq. (18) rather than simply increasing it arbitrarily. The parental population size $\mu = 12$ is large enough to overcome the huge noise influence of $\sigma_\epsilon^* = 5$ in great distance from the optimizer.

In contrast to the previous case (constant noise strength) the mutation strength dynamics

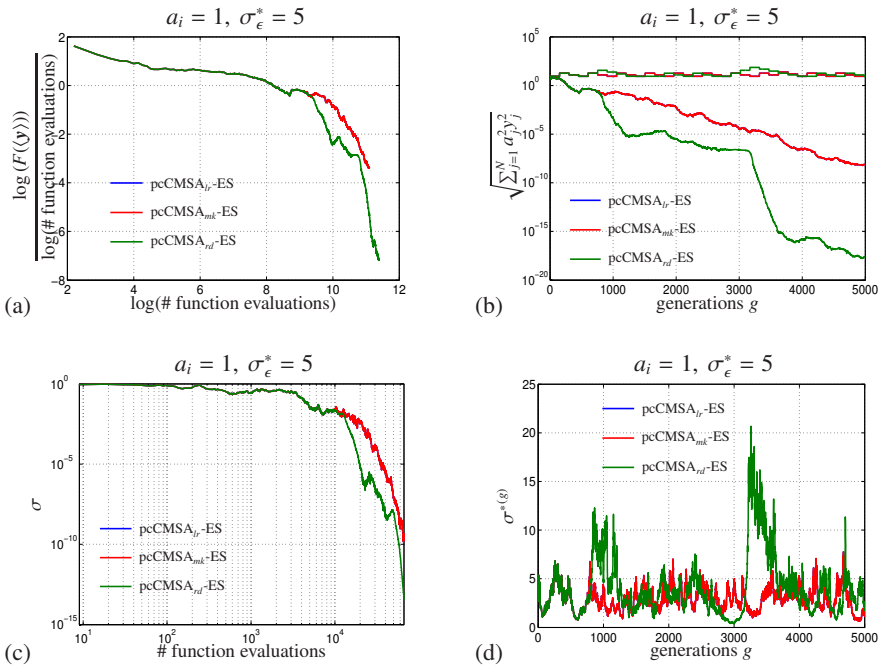

Figure 10: Comparison of the pcCMSA-ES variants on the noisy sphere model with noise of constant normalized variance $\sigma^*_\epsilon = 5$. The λ dynamics are incorporated in (b) for visualization of the population size adaptation behavior of the algorithms. Regarding the illustrated run, $pcCMSA_{mk}$-ES and $pcCMSA_{lr}$-ES exhibit similar dynamics and the corresponding curves (blue and red) overlap.

in Fig. 10(c) indicate a successive reduction of the noise strength σ. This is due to the decreasing influence of the fitness proportional noise as the strategies approach the optimizer. As a result the corresponding normalized mutation strength dynamics in (d) are subject to rather large fluctuations.

Turning to the ellipsoid model $a_i = i$ in Fig. 11 the dynamics resulting from pcCMSA$_{lr}$-ES as well as pcCMSA$_{mk}$-ES show no obvious qualitative changes. But in contrast to the sphere model the pcCMSA$_{rd}$-ES using the residual decision method seems not to stop the process of increasing the population size, see (b). An explanation might be the much lower progress established on the ellipsoid model. Because of the slower decent huge fluctuations bias the decision within the residual decision mechanism. That is, the detection routine repeatedly identifies significant stagnations. Conclusively, the pcCMSA$_{rd}$-ES continuously raises the population size. This leads to a smoother long-term decline of the $\sqrt{\sum_{j=1}^{N} a_j^2 y_j^2}$ dynamics in (b). However, it could also yield in wasting function evaluations since pcCMSA$_{lr}$-ES and pcCMSA$_{mk}$-ES obtain comparable results using smaller populations. The corresponding mutations strength dynamics on the ellipsoid model show no apparent qualitative deviations from those obtained on the sphere model, cf. Fig. 10.

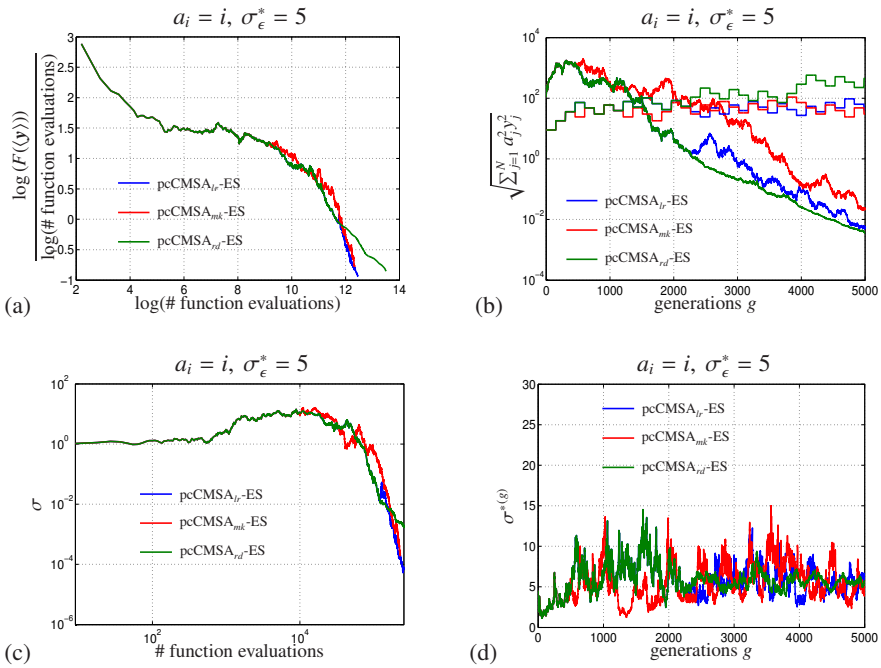

Figure 11: Comparison of the pcCMSA-ES variants on the ellipsoid model being subject to noise of constant normalized variance $\sigma^*_\epsilon = 5$. The λ dynamics are incorporated in (b) for visualization of the population size adaptation behavior of the algorithms.

6.3 Actuator noise

Focusing on actuator noise, the respective dynamics are displayed in Fig. 12 and Fig. 13. On the noisy sphere as well as on the noisy ellipsoid model with $a_i = i$ they qualitatively resemble the constant noise variance case in Sec. 6.1. With growing number of function evaluations the fitness dynamics in (a) show a continuous decline. From (b) we infer that the strategies do recognize stagnations within the fitness dynamics. These stagnations correspond to fluctuations around a certain residual distance which is determined by the actual population sizes and the noise strength. Concerning actuator noise the residual distance is presented in Eq. (21) and is represented within the figures by the colored step functions. Accordingly, the parental centroid's distance to the optimizer is steadily reduced as the algorithm variants increase the population sizes. Alike the case of additive fitness noise of constant variance the mutation strength dynamics approach a steady state. In (d) the corresponding normalized quantities are ascending.

On the ellipsoid model pcCMSA$_{lr}$-ES and pcCMSA$_{mk}$-ES overlap again. This is due to the similar nature of their inner detection mechanism which investigates the slope of the underlying trend. The pcCMSA$_{rd}$-ES on the other hand does not show a quantifiable better performance. Again the original CMSA-ES in Fig. 1 would only approach the residual distance governed by the initial population sizes ($\mu = 3, \lambda = 9$). That is, the three pcCMSA-ES algorithms generate candidate solutions in much closer vicinity to the optimizer. Thus the choice of the detection method within the pcCMSA-ES is not substantial to generate beneficial solutions in the presents of actuator noise on the ellipsoid model.

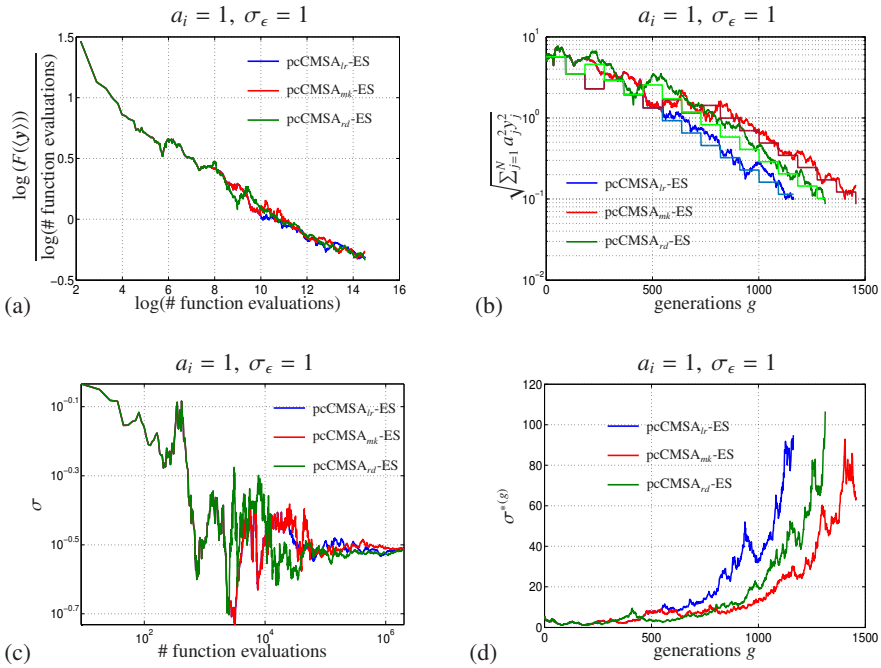

Figure 12: Comparison of the pcCMSA-ES variants on the noisy sphere model applying actuator noise of variance $\sigma_\epsilon = 1$.

7 Discussion

In this paper, an algorithm for the treatment of noisy optimization problems has been developed. The algorithm is based on the well-known CMSA evolution strategy which represents one of the state-of-the-art evolutionary algorithms. Within its concept we integrated mechanisms for the recognition of stagnations. Stagnations in the strategy's progress indicate the presence of noise. In addition to identifying noise related stagnations the developed algorithm also applies appropriate countermeasures. To ensure the discovery of improved candidate solution the algorithm increases the size of the parental as well as the offspring population as soon as stagnations are found. If no further stagnations are discovered in subsequent tests the population size is slowly reduced again to avoid unnecessary function evaluations. This way the algorithm is capable to adapt the appropriate populations size. Accordingly, the adjusted CMSA-ES algorithm is denoted population control covariance matrix self-adaptation evolution strategy, briefly pcCMSA-ES.

Stagnation has been identified within the pcCMSA-ES in three different ways. On the one hand it becomes noticeable by the absence of a clearly negative trends within the noisy fitness dynamics. Therefore, the slope of the respective trend can be determined in two ways. The first method analyzes the fitness dynamics searching for a negative linear trend. That is, the trend is estimated to a predefined confidence level by use of linear regression analysis. Integrating this method in the pcCMSA-ES it is referred to as pcCMSA$_{lr}$-ES. The second method is denoted pcCMSA$_{mk}$-ES. It is based on the Mann-Kendall test which relies on the relative ranking of the analyzed data. The Mann-Kendall test statistically assess if a monotonic downward trend exists within the observations. Due to the proximity of the

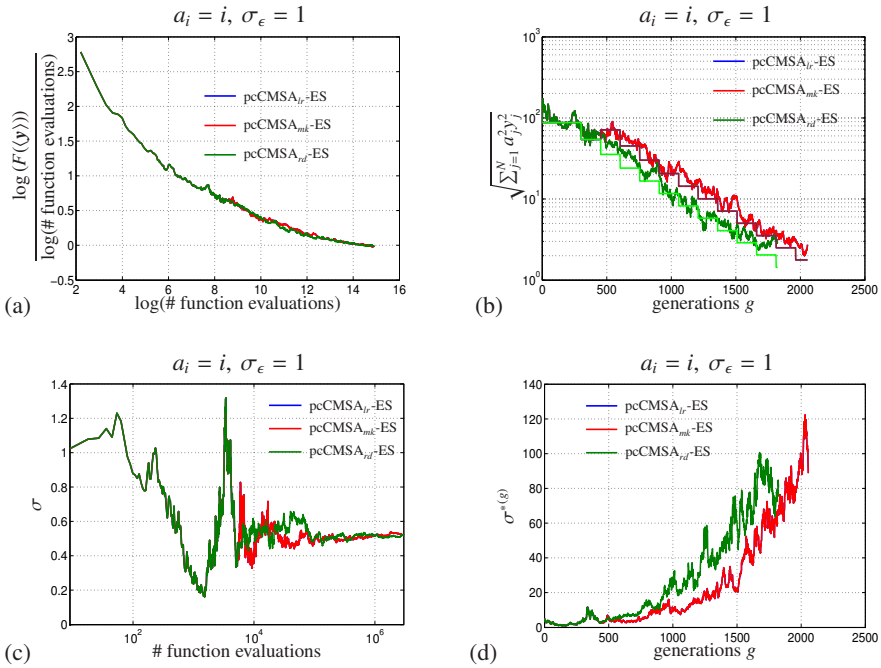

Figure 13: Comparison of the pcCMSA-ES variants on the noisy ellipsoid model with actuator noise of variance $\sigma_\epsilon = 1$.

Mann-Kendall tests to regression analysis, both methods yield very similar results. On the other hand we investigate a third method, which separates the fitness dynamics into their trend and residual components. The residuals are then tested for autocorrelation by use of the Ljung-Box hypothesis test. We suppose that the presence of significant autocorrelations within the residual series indicates ongoing progress, i.e. a continuing trend within the fitness values. Conversely, we interpret the absence of autocorrelation as follows. The residuals appear independent and identically distributed around the associated trend components. This behavior is particularly observed when the fitness values fluctuate around a stable attractor. Thus the slope of the trend is approximately zero and the fitness dynamics are stationary. We refer to this method as residual decision and the associated algorithm as pcCMSA$_{rd}$-ES.

We have tested the algorithm variants on the noisy ellipsoid model. The noise is modeled in three different ways. As additive fitness noise with constant variance σ_ϵ and with constant normalized variance σ_ϵ^*, respectively. As well as by use of the actuator noise model. Conclusively, the pcCMSA-ES represents an advantageous procedure to improve the outcome of the noisy optimization problem. Regardless of the detection method algorithm 7 is able to detect noise influence on the fitness values. Increasing the population sizes the algorithm realizes further progress in direction of the optimizer.

The strategies require the determination of additional strategy parameters. This gives rise to additional investigation concerning the optimal configuration of the pcCMSA-ES. Regardless of the detection mechanism the interval length L of the fitness values considered in a single test decision has to be examined more closely. Within the previous simulations an interval of length $L = 3N$ was considered on the sphere model and $L = 5N$ on the ellipsoid, respectively. The parameter L is connected to the search space dimension N. This choice

has to be regarded as an initial guess to realize an efficient optimization procedure. However, a proper configuration rule still has to be developed since L has significant influence on the identification of fitness stagnations. For the linear regression approach L has to be sufficiently large to ensure the applicability of the central limit theorem. Also the Mann-Kendall test need a sufficiently larger sample size to guarantee reliable test decisions. Especially within the pcCMSA$_{rd}$-ES the parameter L has a great influence on the decomposition process since it determines the estimation quality of the moving average filter. The interval length L defines the length of the residual series r_t which is tested for autocorrelation. It has to be sufficiently long in order to provide reliable test results, see Sec. 4.3.2.

Additionally, the parameter L effects the lead time of the algorithm needed to establish an initial interval of fitness observations $\mathcal{F}_{int}$ as well as the waiting time *wait*. The parameter *wait* governs the length of the waiting period after a single population adjustment. After a transient phase of *wait* generations the algorithm starts again with the analysis of the fitness dynamics. It is not evident if the parameter *wait* needs to be depending on the length L of the fitness interval. The waiting time is essential to prevent wrong test decisions based on fitness dynamics resulting from different population specifications. A beneficial parameter setting has to be determined in future empirical investigations.

Another parameter to be determined is the hypothesis test significance level α. In our simulations it was set to $\alpha = 0.05$. Other choices of α can potentially be used to strengthen the reliability of test decisions. At last the population adaptation parameter c_μ has to be set appropriately. It controls the ascent of the population size and by implication the number of function evaluations as well as the strategy's approach to the optimizer. Using large values of c_μ reduces the residual distance rapidly, but on the other hand it might cause the strategy to consume too much function evaluations at a time where smaller populations still would realize a good performance. Also the factor by which the populations are reduced again when no obvious stagnations are present has to be determined in further studies. Our first choice $\max\left(\mu_{\min}, \lfloor \mu/(1+\frac{1}{c_\mu}) \rfloor\right)$ works well for $c_\mu = 2$ on the noisy ellipsoid model. But considering different values of c_μ effects this reduction factor and might impair the adaptation process of the pcCMSA-ES.

Conclusively, empirical investigations concerning the appropriate strategy parameter setting appear to be the next step. However, the application of the pcCMSA-ES to other noisy fitness environments in order to validate its suitability might be interesting.

Acknowledgements

This work was supported by the Austrian Science Fund FWF under grant P22649-N23.

References

[1] D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer, 2002.

[2] D. V. Arnold and H.-G. Beyer. Local Performance of the $(\mu/\mu_I, \lambda)$-ES in a Noisy Environment. In M. W. and S. W., editors, *Foundations of Genetic Algorithms VI*, pages 127–141. Morgan Kaufmann, 2001.

[3] S. Astete-Morales, M.-L. Cauwet, and O. Teytaud. Evolution strategies with additive noise: A convergence rate lower bound. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, FOGA '15, pages 76–84, New York, NY, USA, 2015. ACM.

[4] H.-G. Beyer. *The Theory of Evolution Strategies*. Natural Computing Series. Springer, Heidelberg, 2001.

[5] H.-G. Beyer, D. V. Arnold, and S. Meyer-Nieberg. A New Approach for Predicting the Final Outcome of Evolution Strategy Optimization Under Noise. *Genetic Programming and Evolvable Machines*, 6(1):7–24, 2005.

[6] H.-G. Beyer and M. Hellwig. The Dynamics of Cumulative Step-Size Adaptation on the Ellipsoid Model. *Evolutionary Computation*, ():1–1, 2014. doi: 10.1162/EVCO_a_00142.

[7] H.-G. Beyer and A. Melkozerov. The Dynamics of Self-Adaptive Multi-Recombinant Evolution Strategies on the General Ellipsoid Model. *IEEE Transactions on Evolutionary Computation*, 18(5):764–778, oct 2014.

[8] H.-G. Beyer, M. Olhofer, and B. Sendhoff. On the Impact of Systematic Noise on the Evolutionary Optimization Performance-A Sphere Model Analysis. *Genetic Programming and Evolvable Machines*, 5(4):327–360, 2004.

[9] H.-G. Beyer and B. Sendhoff. Evolution Strategies for Robust Optimization. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1346–1353, 2006.

[10] P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting*. Number Bd. 1 in Introduction to Time Series and Forecasting. Springer, 2002.

[11] R. O. Gilbert. *Statistical methods for environmental pollution monitoring*. John Wiley & Sons, 1987.

[12] N. Hansen, A. S. P. Niederberger, L. Guzzella, and P. Koumoutsakos. A Method for Handling Uncertainty in Evolutionary Optimization With an Application to Feedback Control of Combustion. *IEEE Trans. Evolutionary Computation*, 13(1):180–197, 2009.

[13] Y. Jin and J. Branke. Evolutionary optimization in uncertain environments-a survey. *IEEE Trans. Evolutionary Computation*, 9(3):303–317, 2005.

[14] M. Kendall. *Rank correlation methods*. C. Griffin, 1948.

[15] J. Kenney. *Mathematics of Statistics*. Number 1-2. Literary Licensing, LLC, 2013.

[16] H. B. Mann. Nonparametric tests against trend. *Econometrica*, 13(3):pp. 245–259, 1945.

[17] A. Melkozerov and H.-G. Beyer. Towards an Analysis of Self-Adaptive Evolution Strategies on the Noisy Ellipsoid Model: Progress Rate and Self-Adaptation Response. In *GECCO 2015, Madrid, Spain*, pages p–p. ACM, 2015.

[18] S. Meyer-Nieberg. *Self-adaptation in evolution strategies*. PhD thesis, Dortmund University of Technology, 2007.

FHV FORSCHUNG

Fachhochschule Vorarlberg
Forschungszentrum
Prozess- und Produkt-Engineering
Hochschulstraße 1
A-6850 Dornbirn

T +43 5572 792 7100
F +43 5572 792 9510
www.fhv.at/forschung/ppe