



# Can Evolution Strategies Benefit from Shrinkage Estimators?

Silja Meyer-Nieberg<sup>(✉)</sup> and Erik Kropat

Department of Computer Science, Universität der Bundeswehr München,  
Werner-Heisenberg Weg 37, 85577 Neubiberg, Germany  
{silja.meyer-nieberg,erik.kropat}@unibw.de

**Abstract.** Evolution strategies are evolutionary algorithms usually applied for solving continuous optimization tasks. As they rely on mutation as one of the main search operators, the control and the adaptation of this process is of high importance. This paper discusses the covariance matrix adaptation in evolution strategies, a central and essential mechanism for the search. The current form bases the estimation of the covariance matrix on small samples sizes compared to the search space dimension which is known to be problematic. This leads to the question, whether the performance of the evolutionary algorithms could be improved if other estimators were utilized. In statistics, several alternative approaches have been considered. Up to now, they have only been seldom applied in evolutionary computation. The paper investigates whether evolution strategies may benefit from linear shrinkage estimators. Several shrinkage targets are considered, integrated in the so-called CMSA-ES, and analyzed experimentally with a special focus on the shrinkage intensity.

## 1 Introduction

Evolutionary algorithms (EAs) are metaheuristics which are based on the principles of natural evolution. As such, they use recombination and mutation to create new search points and perform selection in order to determine which points may pass on their characteristics to succeeding generations. Research in EAs has a long tradition: The first evolutionary algorithms have been introduced in the 1960s. Today, they represent one of the major classes in natural computation. Five main groups exist: genetic algorithms, genetic programming, evolutionary programming, differential evolution, and evolution strategies. The present article focuses on the last class, *evolution strategies* (ESs), which date back to the 70s and are – at least today – mainly applied for continuous optimization. In this area, they have been established as well-performing black-box optimization methods, see e.g. [15]. Their main search operator is mutation in contrast to genetic algorithms which favor recombination.

Evolution strategies, see for example [1, 3, 32] for an introduction, operate with a multivariate normal distribution to generate new search points. The main parameters of the search distribution, the mean  $\mathbf{m}$  and the covariance matrix

$\sigma^2\mathbf{C}$ , must be adapted during the run of an evolution strategy. If they remain constant, the ES may exhibit either only a slow convergence behavior or it may fail in its optimization task entirely. In order to ensure that the mutation control parameters are suited to the current fitness landscape of the function to be optimized, the scale as well as the directions of the mutations must be adapted. For this reason, research in evolution strategies often centers on control techniques for the mutation process with a strong focus on the covariance matrix. Nearly all approaches estimate the current covariance based on the sample. Relying on the sample covariance matrix leads to an ill-posed estimation task in the case of evolution strategies, however: Due to efficiency, the sample size is typically small w.r.t. the search space dimension. This results in a well-known problem in statistics: The estimate may differ considerably from the underlying true covariance [35, 36]. This may be the reason why nearly all current techniques introduce additional correction or regularization terms for example by falling back to the previous covariance matrix and/or by strengthening certain promising directions. Interestingly, these procedures are reminiscent of *shrinkage estimation* in statistics which represents one technique to cope with a poor sample covariance estimate. These similarities lead to the research question of the present paper: If evolution strategies perform a kind of implicit shrinkage, can they profit from the introduction of explicit shrinkage operators?

So far, shrinkage and other covariance matrix estimators have been applied remarkably seldom in the area of evolutionary computation. A literature review resulted in only two papers aside from our previous approaches: The first by Dong and Yao explored an application in the case of Gaussian estimation of distribution algorithms [8]. They faced the problem that the learning of the covariance matrix during the run lead to non positive definite matrices. For this reason, a shrinkage procedure was applied to “repair” the covariance matrix. The approach was similar to in [20] with the exception of an adaptable shrinkage intensity. More recently, Kramer considered a Ledoit-Wolf-estimator based on [19] for an evolution strategy which does not follow a population-based approach but uses a variant of the single-point elitist (1 + 1)-ES. For this reason, the covariance matrix adaptation has to consider past search points and corrects the estimate with shrinkage [17].

The current analysis extends the work carried out in [27, 28] and augments the investigation conducted in [25, 29] for the case of thresholding estimators. [27, 28] presented the first approaches to apply Ledoit-Wolf shrinkage estimators in evolution strategies. In a proof of concept, the shrinkage estimators were combined with an approach stemming from a maximum entropy covariance selection principle. Here, the work is extended by considering several mixture matrices, targets, and choices for the shrinkage intensity which are compared to the original version of the underlying ES variant, the CMSA-ES [5], for noise-free and for noisy optimization.

The paper is structured as follows: First, a brief introduction into evolution strategies with covariance matrix adaptation is provided. Afterwards, we focus on the problem of estimating high-dimensional covariance matrices.

Several shrinkage targets are introduced and their integration into evolution strategies is described. The strategies are assessed and compared to the original ES version in the experimental sections. The paper ends with the conclusions and an outlook regarding open research points.

## 2 Evolution Strategies: A Short Introduction

Let  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  be a continuous function that allows only the evaluation of the function itself but not the derivation of higher order information. This is the area of black-box optimization, where metaheuristics as evolution strategies and similar approaches are often applied. *Evolution strategies* (ESs) are stochastic optimization methods that usually operate with a sample or population of search points also called *candidate solutions*. They distinguish between a population of  $\mu$  parents and  $\lambda$  offspring. In many applications in continuous search spaces, the parent population is discarded after the offspring have been created. Therefore,  $\lambda > \mu$  is required. Evolution strategies use a multivariate normal distribution with mean  $\mathbf{m}^{(g)}$  and covariance matrix  $(\sigma^{(g)})^2 \mathbf{C}^{(g)}$  to create new search points. The mean represents the recombination process and is obtained as the (weighted) centroid of the parent population. The covariance matrix is updated by following one of the established techniques [5, 12]. Sampling  $\lambda$  times from the normal distribution results in the offspring population

$$\mathbf{x}_l = \mathbf{m}^{(g)} + \sigma^{(g)} \mathcal{N}(0, \mathbf{C}^{(g)}), \quad l = 1, \dots, \lambda. \quad (1)$$

Afterwards, the new search points are evaluated using the function  $f$  to be optimized. Evolution strategies then apply deterministic selection in order to determine the next parent population and chose the  $\mu$  best of the  $\lambda$  offspring.

As stated earlier, the parameters of the normal distribution must be adapted in order to allow progress towards the optimal point. Here, the covariance matrix is of great importance and adaptation techniques have received a lot of attention in work on evolution strategies (see [12, 24]). The investigation in this paper centers on the *covariance matrix self-adaptation evolution strategy* (CMSA-ES) [5]. The CMSA-ES divides the adaptation of the covariance  $(\sigma^{(g)})^2 \mathbf{C}^{(g)}$  into two main procedures: the covariance matrix update for  $\mathbf{C}^{(g)}$  and the adaptation of the scale factor  $\sigma^{(g)}$ . The scale factor is also often called step-size or *mutation strength*. Following established practice in evolution strategies, the matrix  $\mathbf{C}^{(g)}$  will be referred to as the *covariance matrix* in the remainder of the paper.

### 2.1 Adaptation I: Covariance Matrix Update

As it is the case for most techniques, the covariance matrix update of the CMSA-ES is based on the sample estimate, that is, on the  $\mu$  best offspring. Considering only the superior candidate solutions shall introduce a bias towards good search regions. Let  $\mathbf{x}_{m:\lambda}$  denote the  $m$ th best of the  $\lambda$  offspring w.r.t. the fitness and let

$$\mathbf{z}_{m:\lambda}^{(g+1)} := \frac{1}{\sigma^{(g)}} \left( \mathbf{x}_{m:\lambda}^{(g+1)} - \mathbf{m}^{(g)} \right), \quad (2)$$

stand for its normalization. The estimate then reads

$$\mathbf{C}_\mu^{(g+1)} := \sum_{m=1}^{\mu} w_m \mathbf{z}_{m:\lambda}^{(g+1)} (\mathbf{z}_{m:\lambda}^{(g+1)})^T \quad (3)$$

with the weights usually set to  $w_m = 1/\mu$  for the CMSA-ES [5]. Please note that the mean  $\mathbf{m}^{(g)}$  is known, thus, the number of degrees of freedom remains equal to  $\mu$ . The sample covariance is then combined with the old covariance, resulting in

$$\mathbf{C}^{(g+1)} := \left(1 - \frac{1}{c_\tau}\right) \mathbf{C}^{(g)} + \frac{1}{c_\tau} \mathbf{C}_\mu^{(g+1)}. \quad (4)$$

The parameter  $c_\tau \in (0, 1)$ ,

$$c_\tau = 1 + \frac{N(N+1)}{2\mu}, \quad (5)$$

see [5] increases with the search space dimension and decreases with the population size.

## 2.2 Adaptation II: Self-adaptation

In addition to the covariance matrix update, the CMSA-ES applies *self-adaptation* to control the mutation strength  $\sigma^{(g)}$ . Self-Adaptation has been developed by Rechenberg [31] and Schwefel [33]. It takes place at the level of the individuals meaning that each population member operates with a distinct set. The strategy parameters are adapted by using similar evolutionary principles as for the main evolutionary algorithm. In other words, they also undergo recombination and mutation processes. Afterwards, they are used in the mutation of the search space position. The influence on the selection is indirect: Self-adaptation is based on the assumption of a stochastic linkage between good objective values and appropriately tuned strategy parameters.

In the case of the CMSA-ES, the mutation process of the mutation strength is realized with the help of the log-normal distribution following

$$\sigma_l^{(g)} = \sigma^{(g)} \exp(\tau \mathcal{N}(0, 1)). \quad (6)$$

The parameter  $\tau$ , the *learning rate*, should scale with  $1/\sqrt{2N}$  [23]. Self-adaptation with recombination has been shown as robust against noise [2] and is therefore considered in this paper. In this case, the variable  $\sigma^{(g)}$  in (6) is the result of the recombination of the mutation strengths. Here, the same recombination type as previously may be used, that is,  $\sigma^{(g+1)} = \sum w_m \sigma_{m:\lambda}$  with  $\sigma_{m:\lambda}$  standing for the mutation strength associated with the  $m$ th best individual. Figure 1 summarizes the main steps of the covariance matrix self-adaptation evolution strategy.

**Require:**  $\lambda, \mu, \mathbf{C}^{(0)}, \mathbf{m}^{(0)}, \sigma^{(0)}, \tau, c_\tau$

- 1:  $g = 0$
- 2: **while** termination criteria not met **do**
- 3:   **for**  $l = 1$  **to**  $\lambda$  **do**
- 4:      $\sigma_l = \sigma^{(g)} \exp(\tau \mathcal{N}(0, 1))$
- 5:      $\mathbf{x}_l = \mathbf{m}^{(g)} + \sigma_l \mathcal{N}(0, \mathbf{C}^{(g)})$
- 6:      $f_l = f(\mathbf{x}_l)$
- 7:   **end for**
- 8:   Select  $(\mathbf{x}_{1:\lambda}, \sigma_{1:\lambda}), \dots, (\mathbf{x}_{\mu:\lambda}, \sigma_{\mu:\lambda})$  according to their fitness  $f_l$
- 9:    $\mathbf{m}^{(g+1)} = \sum_{m=1}^{\mu} w_m \mathbf{x}_{m:\lambda}$
- 10:    $\sigma^{(g+1)} = \sum_{m=1}^{\mu} w_m \sigma_{m:\lambda}$
- 11:    $\mathbf{z}_{m:\lambda} = \frac{\mathbf{x}_{m:\lambda} - \mathbf{m}^{(g)}}{\sigma^{(g)}}$  for  $m = 1, \dots, \mu$
- 12:    $\mathbf{C}_\mu = \sum_{i=1}^{\mu} w_i \mathbf{z}_i \mathbf{z}_i^\top$
- 13:    $\mathbf{C}^{(g+1)} = (1 - \frac{1}{c_\tau}) \mathbf{C}^{(g)} + \frac{1}{c_\tau} \mathbf{C}_\mu^{(g+1)}$
- 14:    $g = g + 1$
- 15: **end while**

**Fig. 1.** The main steps of a CMSA-ES. Normally, the weights  $w_m$  are set to  $w_m = 1/\mu$  for  $m = 1, \dots, \mu$ .

### 3 Covariance Matrix Estimation: Introducing Shrinkage

Using the population covariance matrix necessitates an appropriate sample size with  $\mu \gg N$  for obtaining a high quality estimator. If this is not the case, estimate and “true” covariance may not agree well, especially in the case of high-dimensional search spaces. Among others, the eigen structure may be significantly distorted, see e.g. [20]. However, in evolution strategies typical recommendations for the population sizing state to use an offspring population size  $\lambda$  of either  $\lambda = \mathcal{O}(\log(N))$  or  $\lambda = \mathcal{O}(N)$  and setting  $\mu = \lceil c\lambda \rceil$  with  $c \in (0, 0.5)$ . Thus, either  $\mu/N \rightarrow c$  or even  $\mu/N \rightarrow 0$  for  $N \rightarrow \infty$  holds which disagrees with the requirement.

As stated above, the estimation of covariance matrices has received a lot of attention in statistics, see e.g. [6, 30, 38] and several techniques have been introduced. This paper focuses on *linear shrinkage estimators* that can be computed comparatively efficiently. Other classes, as e.g. thresholding operators for sparse covariance matrix estimation, are currently considered in separate analyses [25, 29]. Following [20, 35], linear shrinkage approaches are based on

$$\mathbf{S}_{\text{est}}(\rho) = \rho \mathbf{F} + (1 - \rho) \mathbf{C}_\mu \quad (7)$$

with  $\mathbf{F}$  the *target* to correct the estimate provided by the sample covariance  $\mathbf{C}_\mu$ . The parameter  $\rho \in (0, 1)$  is called the *shrinkage intensity*. Equation (7) is used to shrink the eigenvalues of  $\mathbf{C}_\mu$  towards the eigenvalues of  $\mathbf{F}$ . The intensity  $\rho$  should be chosen to minimize the expected error

$$\mathbb{E} \left( \|\mathbf{S}_{\text{est}}(\rho) - \Sigma\|_{\mathbf{F}}^2 \right) \quad (8)$$

with  $\|\cdot\|_F^2$  denoting the squared Frobenius norm given by

$$\|\mathbf{A}\|_F^2 = \frac{1}{N} \text{Tr}[\mathbf{A}\mathbf{A}^T], \quad (9)$$

see [20]. Note the factor  $1/N$  is additionally introduced in [20] to normalize the norm w.r.t. the dimension.

Based on (8) and taking into account that the true covariance is unknown in practice, Ledoit and Wolf were able to obtain an optimal shrinkage intensity for the target  $\mathbf{F} = \text{Tr}(\mathbf{C}_\mu)/N \mathbf{I}$  for general probability distributions.

Several other approaches can be identified in literature. On the one hand, different targets can be considered, see e.g. [11, 18, 34, 39]. Schäfer and Strimmer analyzed among others diagonal matrices with equal and unequal variance or special correlation models [34]. Fisher and Sun also allowed for several targets [11] assuming a multivariate normal distribution. Toulumis relaxed the normality assumption, considered several targets, and provided a new non-parametric family of shrinkage estimators [39]. Other authors introduced different estimators, see e.g. [7] or [6]). Recently, Ledoit and Wolf extended their work to include non-linear shrinkage estimators [21, 22].

A problem arises concerning the complexity of the approaches. The associated optimization problem has to be solved which may be a task of its own especially in the case of non-linear estimates. Since the estimation has to be performed in every generation of the ES, only computationally simple approaches can be taken into account. Therefore, the paper focuses on linear shrinkage with shrinkage targets and intensities taken from [11, 19, 20, 39].

Furthermore, transferring shrinkage estimators to ESs needs to consider the situation in which the estimation occurs since it differs from the assumptions in statistical literature:

- First, the covariance matrix  $\Sigma = \mathbf{C}^{(g)}$  that was used to create the offspring is known.
- Second, the sample is based on truncation selection. Therefore, the variables cannot be assumed to be independent and identically distributed (iid). However, this is one of the main assumptions for deriving alternative estimators for the covariance. In the case of evolution strategies, the sample  $\mathbf{x}_{1:\lambda}, \dots, \mathbf{x}_{\mu:\lambda}$  would only represent normally distributed random variables if there were no selection pressure.

In this context, it is interesting to note that in the discussion [12] with respect to the setting of the CMA-ES control parameters it is argued to choose the values so that the distribution of the random variables would remain the same if there were no selection effects. This paper uses a similar argument to justify the usage of the shrinkage intensities obtained for assuming iid random or even normally distributed random variables. Since we are aware of the fact that the situation may differ considerably from the prerequisites in the statistical literature, other settings are also taken into account.

Before continuing, a closer look at the covariance matrix update (4) may be interesting. Equation (4) of the ES algorithm represents a special case of

shrinkage with the old covariance matrix as the target. The shrinkage intensity is determined with  $c_\tau = 1 + [N(N+1)]/[2\mu]$ , Eq. (5), as  $\rho = 1 - 1/c_\tau$ . If  $\mu \gg N$ ,  $\rho \approx 0$  holds and the sample covariance is only slightly corrected. Otherwise, if  $N \gg \mu$ , the old covariance gains importance. To illustrate the main effects of (4), the eigenspace of  $\mathbf{C}^{(g)}$  is considered. Since the covariance matrix is a positive definite matrix, a spectral composition of  $\mathbf{C}^{(g)}$  with  $\mathbf{C}^{(g)} = \mathbf{M}^\top \Lambda \mathbf{M}$  can be conducted. The modal matrix  $\mathbf{M} = (\mathbf{v}_1, \dots, \mathbf{v}_N)$  contains the eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_N$  of  $\mathbf{C}^{(g)}$ , whereas  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$  represents a diagonal matrix with the corresponding eigenvalues  $\lambda_1, \dots, \lambda_N$ . The representation  $\mathbf{C}_C^{(g+1)}$  of  $\mathbf{C}^{(g+1)}$  in the eigenspace of  $\mathbf{C}^{(g)}$  then reads

$$\begin{aligned} \mathbf{C}_C^{(g+1)} &= \rho \Lambda + (1 - \rho) \mathbf{C}_\mu^C \\ &= \text{diag}(\mathbf{C}_\mu^C) + \rho \left( \Lambda - \text{diag}(\mathbf{C}_\mu^C) \right) + (1 - \rho) \left( \mathbf{L}_\mu^C + \mathbf{U}_\mu^C \right) \end{aligned} \quad (10)$$

with  $\mathbf{C}_\mu^C = \mathbf{M}^\top \mathbf{C}_\mu \mathbf{M}$ . The matrix  $\mathbf{L}_\mu^C$  denotes the matrix with the entries of  $\mathbf{C}_\mu^C$  below the diagonal, whereas  $\mathbf{U}_\mu^C$  comprises the elements above. As Eq. (10) shows, the covariance matrix update (4) decreases the off-diagonal elements of the population covariance in the eigenspace. In the case of the diagonal entries, two cases may appear: if  $c_{\mu_{ii}}^C < \lambda_i$ , the new entry is in the interval  $[c_{\mu_{ii}}^C, \lambda_i]$  and thus the estimate increases towards  $\lambda_i$ , otherwise it is shrunk towards  $\lambda_i$ . Thus, in the eigenspace of  $\mathbf{C}^{(g)}$ , Eq. (10) behaves similarly to shrinkage with a diagonal matrix as target matrix. Therefore, in original space, it shrinks the eigenvalues of the population matrix towards those of the target. In contrast to shrinkage, the target matrix is the old covariance (which is not obtainable in the general case). If the sample were drawn from independent and identically distributed variables, the update would “correct” the distortion due to the small sample size with the previous and in that case also the true covariance. Considering that a shrinkage procedure is already present in the original CMSA-ES, the question naturally arises, whether the strategy may benefit from additional corrections of the sample covariance.

Applying shrinkage requires among others the choice of an appropriate target. Most approaches consider regular structures as e.g. the scaled unity matrix, diagonal matrices, or matrices with constant correlations. However, a shrinkage towards a regular structure w.r.t. the coordinate system of the original search space does not appear as a good choice concerning the optimization of arbitrary functions.

## 4 Evolution Strategies and Shrinkage Estimators

Since we cannot assume that the covariance matrix adaptation would profit from correcting the estimate towards regular structures in every application case, shrinkage in the original search space is not taken into account. Instead, appropriate space transformations are investigated. The resulting ES algorithms will follow the same general principle:

1. A suitable transformation of the coordinate system is conducted.
2. A shrinkage estimation is performed in the transformed space.
3. After re-transformation to the original space, the result is used for the covariance matrix update.

As illustrated in the previous section, the original CMSA-update (4) shrinks the sample covariance towards a diagonal matrix in the eigenspace of the previous covariance matrix  $\mathbf{C}^{(g)}$ . Therefore, this eigenspace will also be taken into account for the current investigation.

Space transformations have been considered in ESs before. For example, Hansen argued in [14] that changing the coordinate system may improve the performance. For this reason, he introduced an adaptive encoding for the CMA-ES. It is based on a spectral decomposition of the covariance matrix. New search points are created in the eigenspace of the covariance matrix. Similar to [14], we assume that the ES may benefit from a change of the coordinate system. However, the covariance matrix adaptation and estimation which in [14] occur in the original space will be carried out in the transformed space.

Furthermore, other spaces may be also be beneficial. For example, [37] introduced an additional potential transformation. In [37], the authors were faced with the task to obtain a reliable covariance matrix. To this end, a sample covariance matrix  $\mathbf{S}_i$  was combined with a pooled variance matrix  $\mathbf{C}_p$  – similar to (4)

$$\mathbf{S}_{mix}(\xi) = \xi \mathbf{C}_p + (1 - \xi) \mathbf{S}_i \tag{11}$$

with the parameter  $\xi$  to be determined. To continue, the authors switched to the eigenspace of the non-weighted mixture matrix where they followed a maximal entropy approach to determine an improved estimate of the covariance matrix. Based on [14,37], this paper considers the following choices for the transformation matrix which arise as combinations of the population covariance matrix  $\mathbf{C}_\mu$  and  $\mathbf{C}^{(g)}$

$$\mathbf{S}_{mix} = \mathbf{C}^{(g)} + \mathbf{C}_\mu, \tag{12}$$

$$\mathbf{S}_{(g+1)} = (1 - c_\tau) \mathbf{C}^{(g)} + c_\tau \mathbf{C}_\mu, \tag{13}$$

$$\mathbf{S}_{(g)} = \mathbf{C}^{(g)}. \tag{14}$$

The variants (12)–(14) are based on different assumptions: The first (12) follows [37]. The influence of the old covariance and the population covariance are balanced. Structural changes caused by  $\mathbf{C}_\mu$  will be dampened but will influence the result more strongly than in the case of (13) and (14). The second (14) uses the covariance mixture that appears in the original CMSA-ES. Depending on the size of  $c_\tau$ , which is in turn a function of  $\mu$  and  $N$ , see (5), the influence of the population covariance matrix may be stronger or lesser. The third considers the eigenspace of the old covariance matrix and reduces therefore the influence of the new estimate. Equations (12)–(14) are used to change the coordinate system. The representations of the covariance matrices in the eigenspace are given as  $\mathbf{C}_\mu^S := \mathbf{M}_S^T \mathbf{C}_\mu \mathbf{M}_S$  and  $\mathbf{C}_S := \mathbf{M}_S^T \mathbf{C}^{(g)} \mathbf{M}_S$  with  $S$  standing for one of the



variants (12)–(14). Assuming that for example (14) is used, the following steps are performed

- spectral decomposition:  $\mathbf{M}, \mathbf{D} \leftarrow \text{spectral}(\mathbf{S}_{(g)})$ ,
- determination of  $\mathbf{C}_\mu^S := \mathbf{M}_S^T \mathbf{C}_\mu \mathbf{M}_S$  and  $\mathbf{C}_S := \mathbf{M}_S^T \mathbf{C}^{(g)} \mathbf{M}_S$ ,
- shrinkage estimation resulting in  $\hat{\mathbf{C}}_{\text{shr}}$ ,
- retransformation  $\mathbf{C}_\mu = \mathbf{M}^T \hat{\mathbf{C}}_{\text{shr}} \mathbf{M}$ ,
- covariance adaptation

$$\mathbf{C}^{(g+1)} = \left(1 - \frac{1}{c_\tau}\right) \mathbf{C}^{(g)} + \frac{1}{c_\tau} \mathbf{C}_\mu.$$

They substitute Line 13 of the CMSA-ES algorithm in Fig. 1. Once the space transformation has been conducted, different targets can be taken into account. The analysis considers the matrices

$$\mathbf{F}_u = v \mathbf{I}, \quad (15)$$

with  $v = \text{Tr}(\mathbf{C}_\mu^S)/N$  [20],

$$\mathbf{F}_d = \text{diag}(\mathbf{C}_\mu^S) \quad (16)$$

the diagonal entries of  $\mathbf{C}_\mu^S$  [11, 39], the constant correlation model with matrix  $\mathbf{F}_c$  the entries of which read

$$f_{ij} = \begin{cases} s_{ii} & \text{if } i = j \\ \bar{r} \sqrt{s_{ii} s_{jj}} & \text{if } i \neq j \end{cases} \quad (17)$$

and  $\bar{r} = 2/((N-1)N) \sum_{i=1}^{N-1} \sum_{j=i+1}^N s_{ij} / \sqrt{s_{ii} s_{jj}}$  [19]. The shrinkage intensities are taken from the corresponding publications. For (15) the parameter is based on [20], for (16) it follows [11, 39] while it is taken from [19] in the case of (17).

Since the new ES uses an explicit shrinkage, the question arises whether the additional term consisting of the old covariance matrix in (4) remains necessary or whether the ES may operate solely with shrinkage. Preliminary investigations indicate that the latter strategies perform worse than the original CMSA-ES. Therefore, the current analysis only considers a combination of both. It should also be noted that similar to the original update, the previous covariance could be used to determine the target instead of the sample. Both cases will be investigated more closely in future research.

## 5 Comparing Shrinkage Approaches: An Experimental Analysis

Experiments were conducted to investigate the shrinkage estimators introduced. First, the question of finding a suitable transformation was addressed. To this end, a comparison of the effects of (12)–(14) was carried out for a combination of (16) and the shrinkage intensity taken from [39]. Preliminary experiments

showed that (14) lead to good results. Therefore, the remaining discussion in this paper is restricted to ESs using a transformation with the old covariance matrix. However, the increased variability provided by (12) and (13) should be considered together with (17) or (15) in further experiments.

The analysis considers ES-algorithms which apply shrinkage estimators as defined in (15) to (17). Aside from the CMSA-ES, we denote the strategies as follows

1. CI-ES: a CMSA-ES using (15) as shrinkage target,
2. CC-ES: the CMSA-ES with the constant correlation model (17),
3. FS-ES: the CMSA-ES which uses (16) and follows [11] to determine the shrinkage intensity,
4. Tou-ES: a CMSA-ES based on (16) which uses the shrinkage intensity of [39].

The approaches were coded in MATLAB. In the case of the CI-ES and the C-ES, we used the estimation source code provided by the authors on their webpage<sup>1</sup>. The implementation of the Tou-ES follows closely the R package<sup>2</sup>. In the analysis presented here, we did not use the oracle shrinkage intensity for the FS-ES as in [26] but applied the optimal estimated value (11) in [11, p. 1913]. Therefore, the results may differ from [26]. This ES also operates with a maximal number of fitness evaluations of  $FE_{\max} = 3 \times 10^5 N$ .

## 5.1 Experimental Set-Up

The parameters for the experiments read as follows. Each experiment uses 15 repeats. The initial population is drawn uniformly from  $[-4, 4]^N$ , whereas the mutation strength is chosen from  $[0.25, 1]$ . The search space dimensions were set to 10 and 20. The maximal number of fitness evaluations is given by  $FE_{\max} = 2 \times 10^5 N$ . All evolution strategies use  $\lambda = \lceil \log(3N) + 8 \rceil$  offspring and  $\mu = \lceil \lambda/4 \rceil$  parents. A run terminates prematurely if the difference between the best value so far and the optimal fitness value  $|f_{\text{best}} - f_{\text{opt}}|$  is below a predefined precision set to  $10^{-8}$ . Furthermore, we introduce a restart mechanism into the ESs so that the search is re-initialised when the search has stagnated for  $10 + \lceil 30N/\lambda \rceil$  generations. Stagnation is determined by measuring the best function values in a generation. If the difference between minimal and maximal values of the sample lies below  $10^{-8}$  for the given time-interval, the ES does not make significant movements anymore and the search is started anew.

The experiments are conducted with the help of the black box optimization benchmarking (BBOB) software framework<sup>3</sup> and the test suite, see [13]. The framework allows the analysis of algorithms and provides means to generate tables and figures of the results together with the corresponding legends.

<sup>1</sup> <http://www.econ.uzh.ch/faculty/wolf/publications.html>.

<sup>2</sup> <http://cran.r-project.org/web/packages/ShrinkCovMat>.

<sup>3</sup> The latest software can be obtained from COmparing Continuous Optimisers: COCO (<http://coco.gforge.inria.fr/>).

This paper considers 24 noise-less functions [9]. They consist of four sub-groups: separable functions (function ids 1–5), functions with low/moderate conditioning (ids 6–9), functions with high conditioning (ids 10–14), and two groups of multimodal functions (ids 15–24).

The experiments use the expected running time (ERT) as performance measure. The ERT is defined as the expected value of the function evaluations ( $f$ -evaluations) the algorithm needs to reach the target value with the required precision for the first time, see [13]. In this paper, the estimate

$$\text{ERT} = \frac{\#(FEs(f_{\text{best}} \geq f_{\text{target}}))}{\#succ} \quad (18)$$

is used, that is, the fitness evaluations  $FEs(f_{\text{best}} \geq f_{\text{target}})$  of each run until the fitness of the best individual is smaller than the target value are summed up and divided by the number of successful runs.

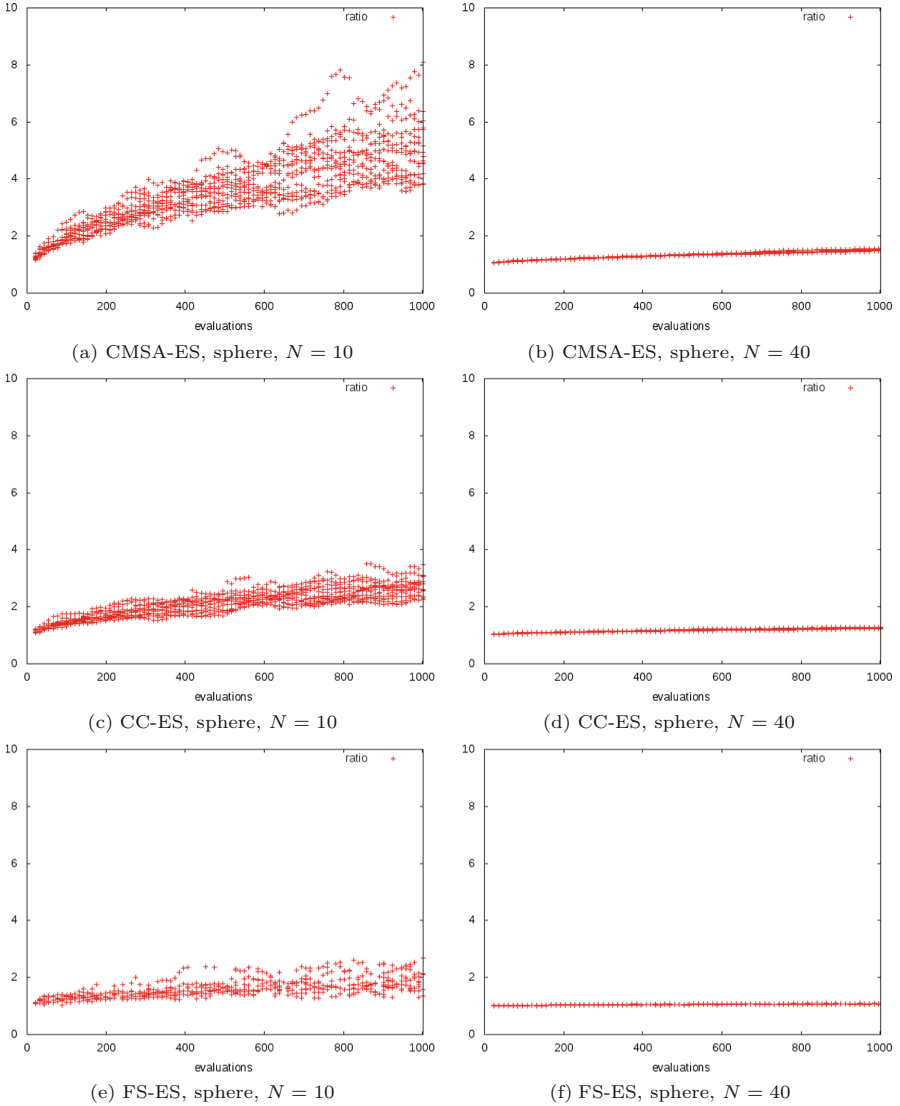
## 5.2 Results and Discussion

First of all, the behavior of the strategies for two exemplary functions, the sphere,  $f(\mathbf{x}) = \|\mathbf{x}\|^2$ , and the discus,  $f(\mathbf{x}) = 10^6 x_1^2 + \sum_{i=2}^N x_i^2$  is investigated. The functions were selected since they represent very different optimization tasks. Figures 2 and 3 show the ratio of the largest to the smallest eigenvalue of the covariance matrix for the CMSA-ES and for two of the shrinkage approaches, the CC-ES and the FS-ES. An ES optimizing the sphere should keep the ratio close to one, whereas the ratio should increase for the discus. In the case of the sphere, the figures illustrate that the largest and smallest eigenvalue develop differently and diverge for  $N = 10$ . Shrinkage causes the problem to be less pronounced.

In the case of the discus, different eigenvalues are expected. All strategies achieve this, the ESs with shrinkage operators show again a lower rate of increase. This may be a hint that the adaptation process of the covariance matrix may be decelerated by the additional shrinkage. Figure 3 shows an even slower increase of the ratio for the FS-ES than for the CC-ES. Whether this lowers the performance is investigated in the second series of experiments. Figure 4 shows the empirical cumulative distribution functions for  $N = 10$  and  $N = 20$ . Whether introducing shrinkage terms improves the performance of the ES depends on the function class. In the case of ill-conditioned functions, shrinkage targets consisting of the diagonal elements of the transformed sample covariance may offer benefits.

The results of the experiments can be examined more closely with the help of Table 1 ( $N = 10$ ) and Table 2 ( $N = 20$ ). They provide the estimate of the expected running time (ERT) for several precision targets ranging from  $10^1$  to  $10^{-7}$ . Also shown is the number of successful runs.

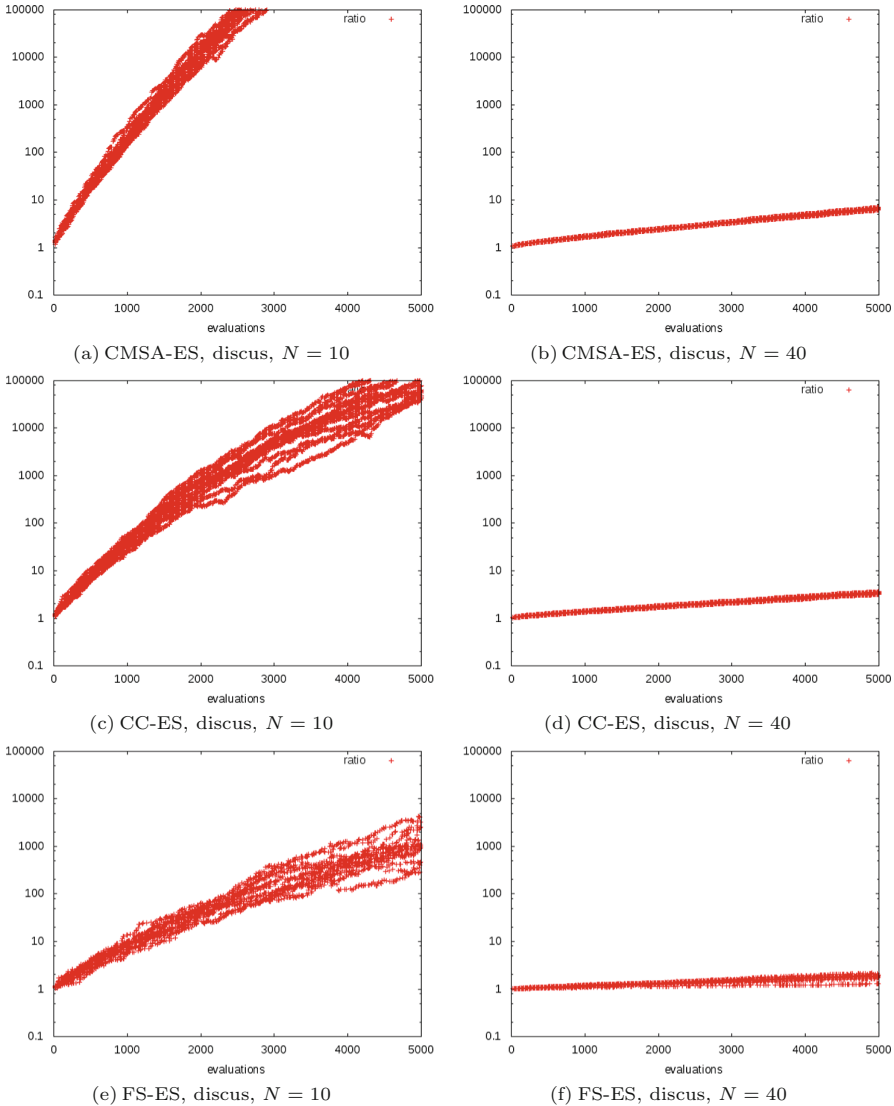
Several functions represent challenges for the ESs considered. These comprise the Rastrigin functions (id 3, id 4, id 15, and id 24) of the test suite, the step ellipsoidal function with a condition number of 100 (id 7), and a multi-modal function with a weak global structure based on Schwefel (id 20). In these cases, all strategies are unable to progress further than the first intermediate precision



**Fig. 2.** The development of the ratio of the largest to the smallest eigenvalue of the covariance on the sphere for the CMSA-ES, the CC-ES, and the FS-ES. Shown are the results from 15 runs per dimensionality.

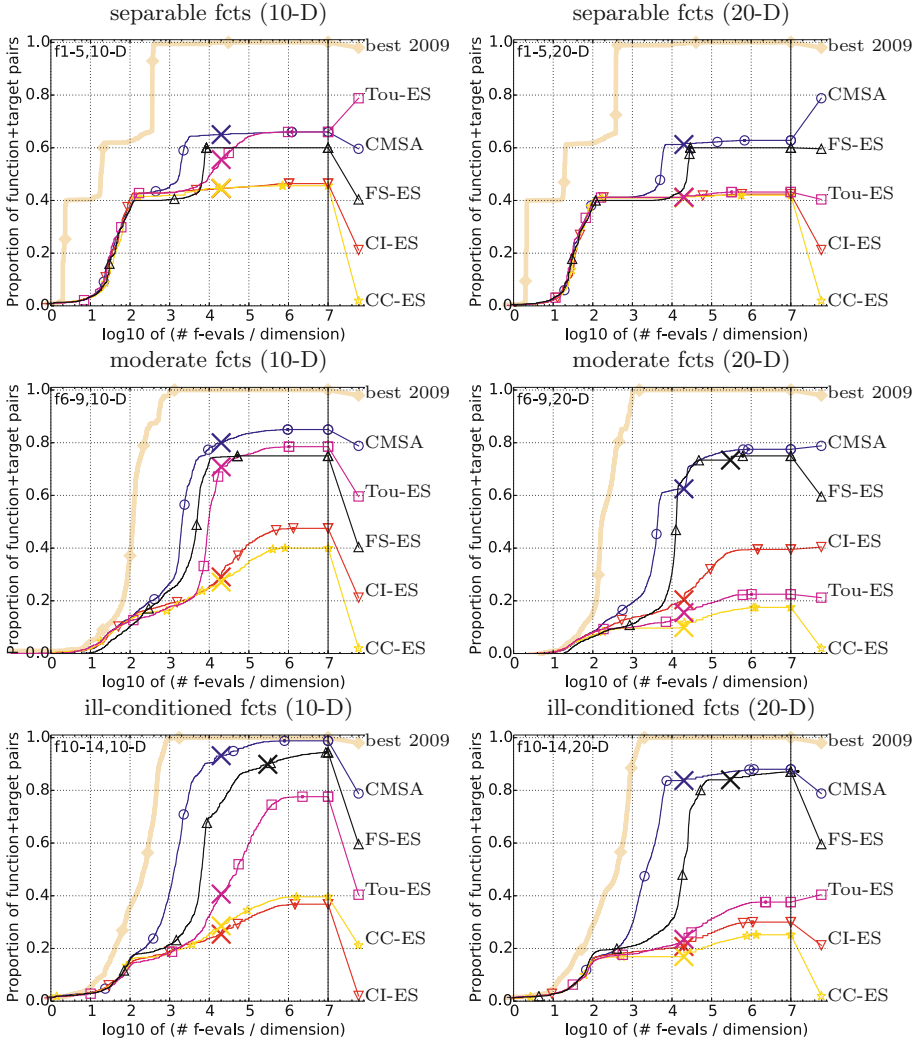
of  $10^1$ . Additionally, for the multi-modal functions, 19 (composite Griewank-Rosenbrock function) and 23 (Katsuura), no strategy could reach  $10^{-1}$ . The functions in question are therefore removed from the tables.

To analyze the remaining functions, the four groups of the test suite are taken into account. The first class comprises the separable functions with id 1 to id 5.



**Fig. 3.** The development of the ratio of the largest to the smallest eigenvalue of the covariance on the discus. Shown are the results from 15 runs per dimensionality.

The three remaining functions, the sphere (f1), the separable ellipsoidal function (f2), and the linear slope (f5) differ in the degree of difficulty for the strategies. All strategies do not show any problems on the sphere or on the slope. Here, several shrinkage variants perform similarly to or sometimes may even surpass the original version. However, the differences are not statistically significant.



**Fig. 4.** Bootstrapped empirical cumulative distribution of the number of objective function evaluations divided by dimension (FEvals/DIM) for 50 targets in  $10^{[-8..2]}$  for 10-D and 20-D.

The more difficult ellipsoidal function, however, cannot be solved by the CC-ES, the CI-ES, and for  $N = 20$  by the Tou-ES. Interestingly, the FS-ES achieves successful runs for all search space dimensionalities. Its performance is significantly worse than the original CMSA-ES, however. Two effects may play a role. The ellipsoidal function, defined by  $f(\mathbf{x}) = \sum_{i=1}^N 10^{6(i-1)/(N-1)} x_i^2$ , is not solved well by ESs with covariance matrices treating all directions with the same weight. The matrix used in the transformation may not be sufficient to provide

the variability required when combined with restrictive structures. The target matrices supplied by the CI-ES may be too regular for the ES to be able to adapt with the necessary velocity. Why the CC-ES also exhibits problems, although it represents a more general model, merits further investigations. The failure of the Tou-ES to achieve the final precision target may be due to the shrinkage intensity since this is the only point where it differs from the FS-ES. For this reason, Sect. 6.2 provides a first analysis of the impact of this parameter on the performance of the ES. In future research, the interaction with the parameter  $c_\tau$  will be investigated more closely. Since this parameter approaches infinity for the typical  $\mu/N$  ratios and increasing dimensionalities, the influence of the sample covariance lessens. Regularizing the covariance matrix may therefore be more important for smaller to medium search space dimensionalities. Concerning the question whether shrinkage improves the performance, no clear answer can be provided for the group of separable functions since the ellipsoidal function apparently requires a faster adaptation than the current versions supply.

The second group of functions consists of the attractive sector function (id 6), the step ellipsoidal function (id 7, results not shown), the original Rosenbrock function (id 8), and a rotated Rosenbrock function (id 9). These functions have low to moderate conditioning. The sector function is difficult to solve for all strategies. For  $N = 20$ , successful runs were recorded for the CMSA-ES and the CI-ES but only for a few cases, i.e., two or three. Therefore, the question arises whether initialization effects may have played a role in other words whether the ES in these cases started in the vicinity of the optimal point. For  $N = 10$  and  $N = 20$ , the FS-ES reaches the final target precision of  $10^{-8}$  in 15 of 15 runs, a result not mirrored by the other strategies. Concerning the Rosenbrock functions (f8 and f9), the CMSA-ES and the FS-ES perform best with the CMSA-ES being superior.

Functions with high conditioning constitute the next group. For the ellipsoidal function (f10), the discus (f11), the bent cigar (f12), the sharp ridge (f13), and the different powers function (f14), the CMSA-ES appears as the best performing strategy, followed by the FS-ES and the Tou-ES. More experiments are required since the bent cigar apparently represents a stronger challenge for the FS-ES for  $N = 10$  than for  $N = 20$ .

In the case of the multi-modal functions (ids 15–24), all ESs encounter problems. Only for the two Gallagher’s problems (id 21 and id 22) successes are recorded. Here, the CMSA-ES and the versions that use the diagonal elements of the sample covariance as shrinkage target show the best results. The FS-ES appears to be a good choice for  $N = 10$  for function 21 whereas the CMSA-ES and the Tou-ES require fewer function evaluations for 22. For  $N = 20$ , only a few runs of all ESs reach the final target precision. Therefore, a comparison for the higher-dimensional search spaces is difficult and is not carried out in this paper.

**Table 1.** Expected running time (ERT in number of function evaluations) divided by the respective best ERT measured during BBOB-2009 in dimension 10. The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and target, the corresponding best ERT in the first row. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries, succeeded by a star, are statistically significantly better (according to the rank-sum test) when compared to all other algorithms of the table, with  $p = 0:05$  or  $p = 10^{-k}$  when the number  $k$  following the star is larger than 1, with Bonferroni correction by the number of instances.

		10-D										
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e-3	1e-5	1e-7	#succ
<b>F1</b>	22	23	23	23	23	23	23	23	23	23	23	15/15
CMSA	4.0(1)	8.6(3)	14(3)	19(5)	26(7)	38(5)	50(7)	50(7)	1041	2602	2954	3338
CC-ES	4.2(2)	9.3(5)	14(3)	19(4)	24(5)	34(6)	43(2)	43(2)	<b>6.2(3)*4</b>	<b>3.2(1)*4</b>	<b>3.4(0.7)*4</b>	<b>3.5(2)*4</b>
CI-ES	3.7(1)	8.4(2)	13(3)	<b>18(4)</b>	<b>23(2)</b>	35(4)	45(6)	45(6)	$\infty$	$\infty$	$\infty$	$\infty$
Tou-ES	4.2(1)	9.1(3)	15(5)	21(4)	27(5)	40(6)	54(8)	54(8)	1.1e4(8256)	150(181)	312(242)	277(374)
FS-ES	3.6(2)	<b>7.8(3)</b>	13(3)	19(4)	25(4)	36(4)	47(4)	47(4)	93(55)	20(2)	19(4)	15(2)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F2</b>	187	190	191	191	193	194	195	195	515	896	1240	1569
CMSA	<b>65(27)*4</b>	<b>85(23)*4</b>	<b>96(20)*4</b>	<b>105(19)*4</b>	<b>109(10)*4</b>	<b>113(19)*4</b>	<b>129(63)*4</b>	<b>129(63)*4</b>	<b>10(21)</b>	<b>13(7)</b>	<b>15(16)*2</b>	<b>16(15)*4</b>
CC-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	29(0.3)	615(725)	2258(2136)	2020(2518)
CI-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	11(17)	37(48)	68(110)	161(315)
Tou-ES	496(794)	850(462)	1517(1236)	1835(1008)	1833(1854)	2248(2160)	2757(929)	2757(929)	76(325)	177(188)	722(1392)	1016(1120)
FS-ES	232(132)	306(96)	342(41)	358(50)	366(81)	376(17)	383(37)	383(37)	5.8(30)	110(101)	167(84)	191(99)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F5</b>	20	20	20	20	20	20	20	20	387	596	797	1014
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F6</b>	20	20	20	20	20	20	20	20	4587	6208	7779	15/15
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F7</b>	20	20	20	20	20	20	20	20	4587	6208	7779	15/15
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F8</b>	20	20	20	20	20	20	20	20	4587	6208	7779	15/15
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F9</b>	20	20	20	20	20	20	20	20	4587	6208	7779	15/15
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)
$\Delta f_{opt}$	l.e.l	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	1e-7	1e0	1e-1	1e-2	1e-3
<b>F10</b>	20	20	20	20	20	20	20	20	4587	6208	7779	15/15
CMSA	<b>12(5)</b>	<b>17(8)</b>	<b>17(7)</b>	<b>17(7)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>17(8)</b>	<b>19(14)</b>	<b>31(30)</b>	<b>58(131)</b>	<b>28(27)</b>
CC-ES	14(9)	19(9)	20(7)	20(8)	20(16)	20(9)	20(15)	20(15)	57(146)	87(157)	229(222)	145(194)
CI-ES	18(15)	102(14)	104(10)	104(16)	104(306)	104(306)	104(9)	104(9)	19(67)	186(268)	610(961)	637(752)
Tou-ES	12(6)	18(6)	19(31)	19(30)	19(16)	19(22)	19(10)	19(10)	59(65)	117(133)	500(750)	307(357)
FS-ES	13(6)	18(5)	19(6)	19(6)	19(9)	19(6)	19(5)	19(5)	<b>6.7(10)</b>	73(90)	167(218)	116(149)

(Continued)



Table 1. (Continued)

$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ	$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ
<b>16</b>	412	623	826	1039	1292	1841	2370	15/15	<b>14</b>	37	98	133	205	392	687	4305	15/15
CMSA	1.4(0.6)	3.3(6)	11(16)	14(21)	19(14)	25(23)	163(178)	6/15	CMSA	1.1(0.7)	2.2(0.9)	2.8(0.7)	3.5(1)	4.8(1)*2	8.7(2)*4	4.8(3)*4	15/15
CC-ES	37(2)	60(87)	183(276)	219(262)	348(346)	452(592)	1206(1139)	1/15	CC-ES	1.2(0.9)	2.4(0.8)	2.9(0.8)	3.6(0.8)	24(16)	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
CI-ES	1.2(0.3)	18(49)	68(132)	193(200)	1088(1159)	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CI-ES	0.98(0.6)	2.4(1)	2.9(2)	3.6(1)	72(42)	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
Tou-ES	5.3(0.3)	89(65)	135(132)	138(92)	151(179)	369(518)	$\infty$ 2 <sup>65</sup>	0/15	Tou-ES	1.4(1)	2.7(1)	3.5(2)	5.4(4)	241(146)	339(453)	2/15	
FS-ES	1.3(0.6)	3.3(1)	5.5(6)	7.5(7)	11(7)	17(19)	28(61)*	15/15	FS-ES	1.2(0.9)	2.6(1)	3.1(0.4)	3.8(0.8)	10(3)	41(10)	73(66)	15/15
$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ	$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ
<b>18</b>	326	921	1114	1267	1315	1343	1570	15/15	<b>18</b>	238	836	7012	15928	27536	37234	42708	15/15
CMSA	3.8(5)	17(13)	18(9)*	18(3)*2	18(9)*2	19(5)*3	19(9)*3	15/15	CMSA	68(234)	129(186)	124(121)	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
CC-ES	75(316)	197(115)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CC-ES	3.3(17)	136(239)	124(171)	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
CI-ES	17(1)	378(279)	784(1350)	2310(4025)	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CI-ES	8.8(10.5)	55(22)	55(22)	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
Tou-ES	11(12)	68(57)	72(43)	76(37)	77(47)	78(33)	79(35)	15/15	Tou-ES	8.5(0.6)	187(239)	211(257)	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
FS-ES	5.5(2)	37(35)	40(24)	42(13)	43(23)	45(28)	46(17)	15/15	FS-ES	1.3(1)	342(575)	334(408)	2822(3673)	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ	$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ
<b>19</b>	200	648	857	993	1065	1138	1185	15/15	<b>19</b>	130	2236	4392	4487	4618	5074	11329	15/15
CMSA	2.5(0.8)	25(17)	24(8)*	22(12)*2	22(10)*3	21(6)*3	21(3)*3	15/15	CMSA	9.5(38)	23(70)	20(48)	20(22)	19(25)	17(38)	7.8(11)	12/15
CC-ES	2.2(1.0)	323(323)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CC-ES	47(235)	109(143)	58(87)	57(77)	55(47)	50(49)	23(47)	8/15
CI-ES	16(3)	212(2236)	212(2236)	86(30)	84(22)	83(18)	83(18)	15/15	CI-ES	33(57)	18(12)	13(10)	17(23)	16(23)	36(63)	35(43)	6/15
Tou-ES	9(9)	91(64)	86(47)	86(30)	84(24)	84(22)	83(18)	15/15	Tou-ES	50(39)	39(134)	32(48)	31(34)	30(24)	28(20)	14(17)	9/15
FS-ES	5.1(6)	37(20)	42(15)	43(9)	43(14)	44(7)	45(20)	15/15	FS-ES	13(12)	14(23)	8.8(10)	8.7(8)	8.5(13)	7.8(9)	3.5(5)	15/15
$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ	$\Delta f_{\text{opt}}$	le1	le0	le-1	le-2	le-3	le-5	le-7	#succ
<b>10</b>	1885	2172	2455	2728	2802	4543	4739	15/15	<b>10</b>	98	2839	6353	6620	6798	8296	10351	6/15
CMSA	6.5(2)*4	7.6(4)*4	7.7(3)*4	7.5(3)*4	7.6(3)*4	4.9(2)*4	4.9(2)*4	15/15	CMSA	25(45)	6.3(6)	13(27)	12(14)	12(13)	10(10)	8.1(28)	13/15
CC-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CC-ES	149(369)	29(73)	24(21)	25(31)	25(21)	30(32)	$\infty$ 2 <sup>65</sup>	0/15
CI-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$ 2 <sup>65</sup>	0/15	CI-ES	48(118)	13(8)	23(32)	49(48)	133(187)	$\infty$	$\infty$ 2 <sup>65</sup>	0/15
Tou-ES	11(225)	301(332)	608(652)	27(4)	27(4)	17(3)	17(1)	15/15	Tou-ES	61(31)	10(16)	17(66)	17(22)	17(25)	6(17)	14(9)	12/15
FS-ES	27(14)	28(11)	29(5)	27(4)	27(4)	17(3)	17(1)	15/15	FS-ES	94(355)	8.8(10)	102(137)	99(186)	96(359)	79(120)	64(168)	15/15

**Table 2.** Expected running time (ERT in number of function evaluations) divided by the respective best ERT measured during BBOB-2009 in dimension 20. The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and target, the corresponding best ERT in the first row. The different target  $\Delta f$ -values are shown in the top row. #succ is the number of trials that reached the (final) target  $f_{\text{opt}} + 10^{-8}$ . The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries, succeeded by a star, are statistically significantly better (according to the rank-sum test) when compared to all other algorithms of the table, with  $p = 0:05$  or  $p = 10^{-k}$  when the number  $k$  following the star is larger than 1, with Bonferroni correction by the number of instances.

		20-D															
$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>F2</b>	385	386	387	388	390	391	393	15/15	<b>F12</b>	1042	1938	2740	3156	4140	12407	13827	15/15
CM-SA	<b>173(38)*</b>	<b>240(32)*</b>	<b>265(27)*</b>	<b>273(25)*</b>	<b>277(44)*</b>	<b>285(42)*</b>	<b>293(24)*</b>	15/15	CM-SA	2.5(4)	10(9)	<b>13(8)</b>	<b>15(8)</b>	<b>14(6)*</b>	<b>5.9(2)*</b>	<b>6.2(2)*</b>	15/15
CC-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CC-ES	60(192)	182(361)	585(949)	1775(2599)	$\infty$	$\infty$	$\infty$	0/15
CI-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CI-ES	97(192)	569(465)	2045(1606)	1775(2567)	$\infty$	$\infty$	$\infty$	0/15
Tou-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	Tou-ES	97(96)	137(242)	315(657)	1775(1204)	$\infty$	$\infty$	$\infty$	0/15
FS-ES	666(283)	987(299)	1172(213)	1235(191)	1286(96)	1307(119)	1320(123)	15/15	FS-ES	<b>1.4(0.1)</b>	49(110)	88(90)	123(126)	131(83)	74(26)	85(18)	15/15
$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>F5</b>	41	41	41	41	41	41	41	15/15	<b>F14</b>	75	239	304	451	932	1648	15661	15/15
CM-SA	12(5)	15(2)	15(6)	15(7)	15(3)	15(6)	15(6)	15/15	CM-SA	1.8(0.7)	1.9(0.6)	2.5(0.3)	3.2(0.4)	<b>5.2(0.6)*</b>	<b>11(4)*</b>	<b>4.2(1)*</b>	15/15
CC-ES	14(6)	18(14)	19(14)	19(7)	19(3)	19(15)	19(15)	15/15	CC-ES	1.8(1)	1.8(0.6)	2.4(0.4)	3.4(0.9)	23(7)	$\infty$	$\infty$	0/15
CI-ES	13(5)	17(5)	18(7)	18(10)	18(7)	18(3)	18(5)	15/15	CI-ES	1.8(0.8)	1.9(0.8)	2.4(0.4)	3.1(0.6)	$\infty$	$\infty$	$\infty$	0/15
Tou-ES	11(3)	14(3)	14(6)	14(4)	14(6)	14(5)	14(5)	15/15	Tou-ES	<b>1.5(0.6)</b>	<b>1.7(0.5)</b>	<b>2.3(0.6)</b>	<b>3.5(0.7)</b>	27(18)	$\infty$	$\infty$	0/15
FS-ES	14(4)	17(15)	18(5)	18(6)	18(6)	18(15)	18(5)	15/15	FS-ES	2.0(0.8)	1.8(0.4)	<b>2.3(0.4)</b>	<b>3.0(0.3)</b>	11(2)	78(22)	32(6)	15/15

(continued)

Table 2. (continued)

$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>f6</b>	1296	2343	3413	4255	5220	6728	8409	15/15	<b>f17</b>	63	1030	4005	12242	30677	56288	80472	15/15
CMSA	1.5(1)	2.5(2)	4.6(3)	12(9)	34(51)	80(45)	331(595)	2/15	CMSA	0.93(0.9)	196(389)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CC-ES	29(19)	88(172)	108(92)	154(283)	144(131)	253(236)	206(198)	3/15	CC-ES	1.1(1.0)	445(1942)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CI-ES	80(234)	471(726)	1643(1319)	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CI-ES	1.3(3)	341(97)	650(849)	$\infty$	$\infty$	$\infty$	$\infty$	4e5
Tou-ES	56(10)	275(430)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	Tou-ES	1.2(0.8)	778(1844)	1399(1398)	$\infty$	$\infty$	$\infty$	$\infty$	4e5
FS-ES	2.0(1.0)	7.6(9)	18(25)	29(30)	42(34)	59(38)	71(32)	15/15	FS-ES	0.95(0.9)	6658(2e4)	9738(2e4)	6862(1e4)	$\infty$	$\infty$	$\infty$	6e6
$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>f8</b>	2039	3871	4040	4148	4219	4371	4484	15/15	<b>f21</b>	561	6541	14103	14318	14643	15567	17589	15/15
CMSA	11(5)*2	30(30)	31(8)*	31(78)*	31(5)*	31(28)*	31(46)*	13/15	CMSA	179(356)	245(183)	113(149)	112(175)	109(82)	103(109)	91(91)	3/15
CC-ES	106(61)	727(1144)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CC-ES	357(713)	398(275)	184(206)	182(168)	178(157)	167(122)	148(267)	2/15
CI-ES	2752(4562)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CI-ES	357(534)	168(199)	113(128)	112(126)	109(75)	109(84)	91(119)	3/15
Tou-ES	96(45)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	Tou-ES	179(356)	122(183)	78(92)	77(70)	75(96)	71(167)	63(131)	4/15
FS-ES	27(13)	39(8)	48(14)	51(13)	53(17)	54(11)	54(15)	15/15	FS-ES	3887(8016)	35963(9403)	2765(3829)	2724(2093)	2663(3483)	2505(2794)	2217(2558)	2/15
$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>f9</b>	1716	3102	3277	3379	3455	3594	3727	15/15	<b>f22</b>	467	5580	23491	24163	24948	26847	1.3e5	12/15
CMSA	17(6)*2	40(130)*	41(63)*	41(4)*	41(60)*	40(60)*	40(54)*	13/15	CMSA	572(1500)	287(108)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CC-ES	158(107)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CC-ES	980(1714)	466(699)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CI-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CI-ES	312(428)	108(233)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
Tou-ES	88(51)	633(773)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	Tou-ES	572(1285)	287(556)	289(145)	232(393)	225(132)	211(283)	43(66)	1/15
FS-ES	38(6)	194(486)	196(4)	197(448)	196(440)	191(5)	186(6)	14/15	FS-ES	1.5e4(2e4)	1.5e4(2e4)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	6e6
$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
<b>f10</b>	7413	8661	10735	13641	14920	17073	17476	15/15	<b>f23</b>	467	5580	23491	24163	24948	26847	1.3e5	12/15
CMSA	10(2)*4	11(4)*4	9.2(2)*4	7.8(1)*4	7.3(1)*4	6.7(0.6)*4	6.8(1)*4	15/15	CMSA	572(1500)	287(108)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CC-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CC-ES	980(1714)	466(699)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
CI-ES	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	CI-ES	312(428)	108(233)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	4e5
Tou-ES	796(1389)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	0/15	Tou-ES	572(1285)	287(556)	289(145)	232(393)	225(132)	211(283)	43(66)	1/15
FS-ES	36(12)	47(9)	44(3)	37(3)	34(3)	31(2)	31(1)	15/15	FS-ES	1.5e4(2e4)	1.5e4(2e4)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	6e6

## 6 Further Experimental Analyses

The experiments revealed that at least in the eigenspace of the previous covariance matrix and for the intensities considered, the ES does not benefit at all from shrinkage towards more complex structures than offered by a diagonal matrix. This is surprising, since the model with constant correlations would allow more degrees of freedom for the evolutionary process. Therefore, further investigations – concerning especially the question of the choice of the shrinkage intensity – will be carried out. In addition, ESs with the scaled unity matrix as a target do not perform as well as strategies with different diagonal elements. This can possibly be traced back to the fact that this target is the most restrictive and impairs the general adaptability of the ES.

Therefore, the further discussion in this paper is limited to the shrinkage estimators that use the diagonal entries of the sample covariance as correction term. The two variants taken into account, the Fisher-Sun (FS-ES) estimator and the Toulumis technique (Tou-ES) operate with the same target, but the performance of the associated evolution strategies differs. Since the only difference between these two estimators lies in the shrinkage intensity, this section carries out an investigation of the dependency of the performance of the ES on the choice of the factor.

So far, the experiments were restricted to the noise-free case assuming the possibility of exact function evaluations. The effects of shrinkage estimators concerning noisy optimization remain to be taken into account. Since this represents an important research and application area of evolution strategies, exemplary experiments are carried out.

### 6.1 A Brief Investigation Concerning Noisy Optimization

This subsection addresses the question whether shrinkage estimators may be useful in the case of noisy optimization. *Noise* or more generally *uncertainty* is a common and important problem in practical optimization. Following [4, 16], uncertainty comprises noise, robustness, dynamical changes, and function approximations. Many causes exist: Physical measurements may be necessary during the optimization which are usually imprecise to a certain degree. Situational changes may occur – for instance when trying to find the fastest route in a traffic network. For a closer explanation, let us reconsider the function  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  that is to be optimized. Noise means that the function evaluations are not exact and that disturbances, i.e., measurement errors or similar, need to be taken into account. These can be modelled by a random variable  $\epsilon$ . Instead of the exact  $f$ -value at  $\mathbf{x}$  only the noisy  $\tilde{f}(\mathbf{x}) = g(f, \mathbf{x}, \epsilon)$  can be observed. Following the test suite [10], the multiplicative noise model

$$\tilde{f}(\mathbf{x}) = f(\mathbf{x})\epsilon \tag{19}$$

is considered in the paper. In the experiments conducted for the sphere and the Rosenbrock function, the noise term  $\epsilon$  is represented by a log-normally

**Table 3.** ERT and half-interquartile range (90%–10%) divided by the best ERT measured during BBOB 2009 for different  $\Delta f$  values in 10-D.

10-D						
$\Delta f_{\text{opt}}$	10	0.1	1e-3	1e-5	1e-7	#succ
<b>f<sub>101</sub></b>	26	181	194	210	226	15/15
1:CMSA-ES	3.0 <sub>(2)</sub>	1.7 <sub>(0.2)</sub>	2.9 <sub>(0.4)</sub>	3.9 <sub>(0.4)</sub>	4.9 <sub>(0.8)</sub>	15/15
2:FS-ES	2.8 <sub>(1)</sub>	1.6 <sub>(0.4)</sub>	2.8 <sub>(0.2)</sub>	3.6 <sub>(0.5)</sub>	4.5 <sub>(0.3)</sub>	15/15
<b>f<sub>104</sub></b>	610	16641	19364	20764	22011	15/15
1:CMSA-ES	368 <sub>(2)</sub>	140 <sub>(38)</sub>	$\infty$	$\infty$	$\infty$ 3.0e6	0 / 15
2:FS-ES	22 <sub>(27)</sub>	46 <sub>(8)</sub>	<b>46</b> <sub>(79)</sub> *	<b>44</b> <sub>(4)</sub> *	<b>50</b> <sub>(78)</sub> *	13/15
<b>f<sub>107</sub></b>	945	3871	7352	11340	14303	15/15
1:CMSA-ES	1.5 <sub>(2)</sub>	365 <sub>(213)</sub>	224 <sub>(3)</sub>	147 <sub>(4)</sub>	117 <sub>(4)</sub>	15/15
2:FS-ES	0.92 <sub>(0.3)</sub>	<b>2.1</b> <sub>(1)</sub> * <sup>2</sup>	<b>2.7</b> <sub>(1)</sub> * <sup>4</sup>	<b>2.8</b> <sub>(2)</sub> * <sup>4</sup>	<b>2.9</b> <sub>(1)</sub> * <sup>4</sup>	15/15
<b>f<sub>110</sub></b>	11224	7.0e7	$\infty$	$\infty$	$\infty$	0
1:CMSA-ES	89 <sub>(76)</sub>	0.62 <sub>(0.6)</sub>	$\infty$	$\infty$	$\infty$	0/15
2:FS-ES	2.5 <sub>(0.9)</sub>	0.10 <sub>(0.1)</sub>	1.3e7 <sub>(1e7)</sub>	1.3e7 <sub>(1e7)</sub>	1.3e7 <sub>(3e7)</sub>	3/15

distributed random variable  $\exp(\beta\mathcal{N}(0,1))$  [10]. Two variants are taken into account, moderate noise with  $\beta = 0.01$  (id 101 (sphere), id 104 (Rosenbrock)) and severe noise  $\beta = 1$  (ids 107 (sphere) and 110).

For the comparison, the FS-ES is chosen as the representative for the ESs with additional shrinkage. Table 3 shows a promising finding: While both strategies behave comparable on the sphere if the noise does not have a strong effect, the FS-ES appears as superior if the fitness evaluations are severely disturbed. Here, the additional shrinkage may serve to stabilize the covariance matrix update and may thus improve the performance of evolution strategies in the case of noisy optimization. A similar behavior can be observed for the Rosenbrock function. Here, the FS-ES performs better even for low noise levels.

## 6.2 The Role of the Shrinkage Intensity: A Closer Look at the Fisher-Sun Estimate

The preceding sections showed that the shrinkage intensity may have an influence on the performance of the ES, considering that the results of the FS-ES and the Tou-ES differ. Therefore, this section takes a closer look at the Fisher-Sun estimator and analyzes different value settings. We start from the original equation in [11, p. 1913]

$$\hat{\lambda} = \frac{\hat{\beta}_D^2 + \hat{\gamma}_D^2}{\hat{\delta}_D^2}. \quad (20)$$

The parameters in (20) are obtained as follows: Let  $\hat{a}_1 = \text{Tr}(\mathbf{C}_\mu^S)/N$  and

$$\hat{a}_2 = \frac{\mu^2}{N(\mu - 1)(\mu + 2)} \left( \text{Tr}(\mathbf{C}_\mu^{S^2}) - \frac{1}{\mu} (\text{Tr}(\mathbf{C}_\mu^S))^2 \right). \quad (21)$$

Denote the matrix with the diagonal elements of  $\mathbf{C}_\mu^S$  as  $\mathbf{D}_C$ . Then with

$$\hat{a}_1^* = \text{Tr}(\mathbf{D}_C)/N \text{ and} \quad (22)$$

$$\hat{a}_2^* = \frac{\mu}{N(\mu + 2)} \text{Tr}(\mathbf{D}_C^2) \quad (23)$$

the parameters read

$$\hat{\beta}_D^2 = \frac{1}{\mu} (\hat{a}_2 + N\hat{a}_1^2), \quad (24)$$

$$\hat{\gamma}_D^2 = -\frac{2}{\mu} \hat{a}_2^*, \quad (25)$$

$$\hat{\delta}_D^2 = \frac{\mu + 1}{\mu} \hat{a}_2 + \frac{N}{\mu} \hat{a}_1^2 - \frac{\mu + 2}{\mu} \hat{a}_2^*. \quad (26)$$

Starting from (20), the shrinkage intensity  $\rho = c\hat{\lambda}$  is varied with factors  $c = 0.01$  (---),  $0.1$  (--),  $0.5$  (-),  $1.5$  (+),  $2.0$  (++) and  $c = 10$  (+++). In the case that the resulting parameter exceeds one, it is reset to  $\rho = 1$ . The investigation considered two separable functions with ids 1 (sphere) and 2 (ellipsoidal), the two Rosenbrock functions (id 8, 9) as representatives for lowly and moderately conditioned functions, and aside from the sharp ridge (id 13) the set of functions with high conditioning (ids 10–14).

As it can be inferred from Table 4, the effects of the scaling factor  $\rho$  differ with the function that is to be optimized. For nearly all functions considered a decrease and therefore a smaller influence of the diagonal shrinkage target appears as beneficial. Comparing the results with those in Table 1 shows that the performance depends very strongly on the size of the factor and can be improved decisively. However, it appears difficult to obtain a general recommendation. In some cases, e.g. the discus, id 11, a slight decrease may be preferable whereas the ES in the case of the original Rosenbrock, id 8, apparently operates better with only a small correction by the target. For the set of functions considered, a decrease by half or the factor 0.1 appears to improve the results. However, further investigations – especially into potential adaptation mechanisms of the shrinkage intensity – appear necessary. This is underlined by the findings in the case of the ellipsoidal with id 2 which differ strongly from the rest of the functions. Here, an increase of the intensity improves the performance, albeit not significantly. This may be a hint that a strong influence of the target is preferable. As it can be seen, the “optimal” shrinkage intensity from literature may not represent the best choice for the purposes of optimization. This was to be expected since the situation differs from the iid or even normally distributed case considered in the papers.

**Table 4.** Expected running time (ERT in number of function evaluations) divided by the respective best ERT measured during BBOB-2009 in dimension 10. The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and target, the corresponding best ERT in the first row. The different target  $\Delta f$ -values are shown in the top row. #succ is the number of trials that reached the (final) target  $f_{\text{opt}} + 10^{-8}$ . The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries, succeeded by a star, are statistically significantly better (according to the rank-sum test) when compared to all other algorithms of the table, with  $p = 0:05$  or  $p = 10^{-k}$  when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

		10-D																
		$\Delta f_{\text{opt}}$	1e0	1e1	1e-2	1e-3	1e-5	1e-7	#succ	$\Delta f_{\text{opt}}$	1e1	1e0	1e-1	1e-2	1e-3	1e-5	1e-7	#succ
+	f1	22	23	23	23	23	23	23	15/15	f10	1835	2172	2455	2728	2802	4543	4739	15/15
+	+	4.0(4)	8.8(2)	14(4)	20(4)	24(4)	35(4)	47(4)	15/15	++	817(409)	3460(2763)	1.8e4(2e4)	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	3.4(2)	8.0(1)	13(3)	18(3)	24(4)	35(4)	46(5)	15/15	++	716(731)	4335(4582)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	3.6(2)	8.6(2)	14(7)	19(2)	24(8)	35(4)	47(5)	15/15	++	203(56)	291(224)	358(156)	374(104)	394(80)	296(65)	$\infty$	3/6/6
-	-	3.6(2)	7.8(2)	13(4)	19(2)	24(2)	35(3)	48(3)	15/15	-	12(5)	14(3)	14(1)	13(2)	13(2)	8.1(1)	8.0(1)	15/15
-	-	4.6(1)	10(4)	15(4)	21(4)	27(3)	38(7)	49(10)	15/15	-	10(4)	13(4)	13(2)	12(1)	12(2)	7.7(1)	7.6(1)	15/15
-	-	4.3(4)	10(2)	15(3)	21(2)	27(6)	39(6)	51(6)	15/15	-	11(3)	13(3)	12(2)	12(3)	12(2)	7.7(1.0)	7.7(1)	15/15
$\Delta f_{\text{opt}}$	f2	187	190	191	191	193	194	195	15/15	f11	266	1041	2602	2954	3338	4092	4843	15/15
+	+	68(17)	83(19)	88(11)	91(16)	92(14)	94(15)	95(15)	15/15	++	3.6e4(3e4)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	69(19)	83(15)	94(18)	96(11)	97(17)	98(16)	99(11)	15/15	++	2.3e4(3e4)	4.3e4(6e4)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	119(85)	274(277)	365(198)	426(381)	503(364)	685(649)	2204(4038)	15/15	++	1321(905)	711(239)	366(136)	384(109)	414(78)	1096(1755)	$\infty$	3/6/6
-	-	113(37)	148(48)	171(19)	176(21)	178(32)	181(11)	184(24)	15/15	-	30(3)	10(2)	4.8(0.5)	4.5(1)*2	4.3(0.5)*2	3.9(0.5)*	3.5(0.5)*	15/15
-	-	105(58)	147(50)	165(25)	171(31)	173(39)	181(48)	184(34)	15/15	-	30(7)	13(2)	6.4(2)	6.4(2)	6.2(1)	5.6(2)	5.0(2)	15/15
-	-	110(32)	151(32)	167(40)	177(41)	182(49)	192(42)	197(30)	15/15	-	33(5)	13(3)	6.7(1)	7.1(2)	6.9(2)	6.2(2)	5.7(2)	15/15
$\Delta f_{\text{opt}}$	f8	326	921	1114	1217	1267	1315	1343	15/15	f12	515	896	1240	1390	1569	3660	5154	15/15
+	+	5.5(0.7)	113(12)	116(187)	133(165)	158(212)	213(555)	$\infty$ 3/6/6	0/15	++	898(0.2)	6694(6693)	3.4e4(5e4)	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	8.2(2)	111(288)	114(240)	157(9)	171(198)	211(195)	$\infty$ 3/6/6	0/15	++	898(1.456)	4102(5856)	$\infty$	$\infty$	$\infty$	$\infty$	$\infty$	3/6/6
+	+	14(71)	91(15)	108(14)	126(58)	143(203)	171(105)	5638(7094)	5/15	+	1.6(0.3)	922(1587)	2376(2139)	3930(3537)	6554(9527)	$\infty$	$\infty$	0/15
-	-	4.0(7)	23(21)	25(17)	25(13)	25(17)	25(8)	25(8)	15/15	-	3.3(0.3)	12(12)	18(10)	21(0)	22(5)	12(3)	9.3(3)	15/15
-	-	4.8(14)	29(22)	29(12)	28(18)	28(13)	28(14)	28(16)	15/15	-	1.7(0.3)	14(16)	19(11)	20(10)	21(11)	12(6)	9.2(4)	15/15
-	-	6.1(18)	16(6)	19(9)	19(9)	19(6)	20(9)	20(9)	15/15	-	1.7(0.3)	14(16)	19(11)	20(10)	21(11)	12(6)	10(3)	15/15
$\Delta f_{\text{opt}}$	f9	200	648	857	993	1065	1138	1185	15/15	f14	37	98	133	205	392	687	4305	15/15
+	+	27(87)	263(506)	262(423)	342(312)	451(443)	751(274)	$\infty$ 3/6/6	0/15	++	1.6(1.0)	3.1(0.9)	3.1(0.9)	3.6(0.7)	13(9)	$\infty$	$\infty$	3/6/6
+	+	2.6(2)	145(17)	124(17)	272(402)	391(394)	648(90)	$\infty$ 3/6/6	0/15	++	2.1(0)	2.1(0)	2.1(0)	2.7(0.7)	11(4)	1322(1084)	$\infty$	0/15
+	+	2.6(2)	93(19)	124(17)	163(19)	197(129)	234(124)	27(3)	0/15	++	1.5(1.4)	3.0(0.7)	3.0(0.7)	3.7(0.7)	11(4)	16(2)	12(6)	15/15
-	-	2.7(0.3)	32(39)	31(17)	30(8)	29(4)	28(10)	27(4)	15/15	-	1.3(0.8)	2.0(0.7)	2.5(0.6)	3.2(0.4)	6.4(2)	14(3)	8.4(4)	15/15
-	-	3.6(6)	32(39)	30(27)	28(4)	27(21)	27(37)	26(36)	15/15	-	1.5(0.7)	2.4(1)	2.9(0.9)	3.5(0.8)	6.6(0.9)	14(3)	8.4(4)	15/15
-	-	2.5(1)	25(16)	27(22)	26(15)	25(5)	25(11)	25(12)	15/15	-	1.1(0.9)	2.3(0.5)	2.8(0.6)	3.4(0.5)	6.3(2)	16(4)	12(7)	15/15

## 7 Conclusions

Evolution strategies are population-based evolutionary algorithms for continuous optimization. They use a multivariate normal distribution to generate new search points. The parameters of this distribution must be adapted in order to achieve a well performing optimization method. For this reason, step-size and covariance matrix adaptation techniques are important topics in research on evolution strategies. This paper considered the covariance matrix adaptation. Current techniques use the sample covariance, an estimator, which may be of poor quality if the sample size is small. Typically, the estimate is corrected with the help of additional terms. The resulting effect is remarkably similar to shrinkage estimation, a method stemming from statistics. There, shrinkage operators have been introduced in order to improve the quality of the sample covariance by correcting the estimate with the help of a target matrix.

The realization that evolution strategies already perform an implicit shrinkage leads to the research question of the present paper: Would an additional inclusion of shrinkage operators improve the performance of evolution strategies? Applying explicit shrinkage in evolution strategies requires several new tasks to be solved: The choice of the target and the combination weight of target and sample covariance, the shrinkage intensity, are crucial. Since an evolution strategy is used to optimize arbitrary functions with various structures, the approach must remain sufficiently adaptable. To achieve this, we considered a transformation of the original search space to conduct the shrinkage.

The experimental analysis took several shrinkage targets into account using the intensity settings of the original publications. Pending further experiments that shall provide more information regarding the shrinkage intensity which may have interfered with the findings, shrinkage targets in the transformed space that use a diagonal matrix consisting of the different entries of the transformed sample covariance appear as the best choices. While introducing additional shrinkage does not always improve the performance in the noise-less case and at times even impairs it, it may offer a means to lessen the impact of noise.

As the experiments showed, the choice of the shrinkage intensity may have a strong influence on the performance. Since the original covariance matrix adaptation performs a further type of shrinkage which lessens the influence of the sample covariance when the search space dimensionality increases, future research will focus on the shrinkage intensity and its interaction with the covariance matrix adaptation.

## References

1. Bäck, T., Foussette, C., Krause, P.: Contemporary Evolution Strategies. Natural Computing. Springer, Heidelberg (2013). <https://doi.org/10.1007/978-3-642-40137-4>
2. Beyer, H.G., Meyer-Nieberg, S.: Self-adaptation of evolution strategies under noisy fitness evaluations. Genet. Program. Evolvable Mach. **7**(4), 295–328 (2006)



3. Beyer, H.G., Schwefel, H.P.: Evolution strategies: a comprehensive introduction. *Nat. Comput.* **1**(1), 3–52 (2002)
4. Beyer, H.G., Sendhoff, B.: Robust optimization - a comprehensive survey. *Comput. Methods Appl. Mech. Eng.* **196**(33–34), 3190–3218 (2007). <https://doi.org/10.1016/j.cma.2007.03.003>
5. Beyer, H.-G., Sendhoff, B.: Covariance matrix adaptation revisited –the CMSA evolution strategy–. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) *PPSN 2008*. LNCS, vol. 5199, pp. 123–132. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-87700-4\\_13](https://doi.org/10.1007/978-3-540-87700-4_13)
6. Chen, X., Wang, Z., McKeown, M.: Shrinkage-to-tapering estimation of large covariance matrices. *IEEE Trans. Signal Process.* **60**(11), 5640–5656 (2012)
7. Chen, Y., Wiesel, A., Eldar, Y.C., Hero, A.O.: Shrinkage algorithms for MMSE covariance estimation. *IEEE Trans. Signal Process.* **58**(10), 5016–5029 (2010)
8. Dong, W., Yao, X.: Covariance matrix repairing in gaussian based EDAs. In: 2007 IEEE Congress on Evolutionary Computation, CEC 2007, pp. 415–422 (2007)
9. Finck, S., Hansen, N., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2010: presentation of the noiseless functions. Technical report, Institute National de Recherche en Informatique et Automatique, 2009/22 (2010)
10. Finck, S., Hansen, N., Ros, R., Auger, A.: Real-parameter black-box optimization benchmarking 2010: Presentation of the noisy functions. Working paper 2009/21 (2010). <http://coco.lri.fr/downloads/download15.03/bbobdocnoisyfunctions.pdf>
11. Fisher, T.J., Sun, X.: Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comput. Stat. Data Anal.* **55**(5), 1909–1918 (2011). <http://www.sciencedirect.com/science/article/pii/S0167947310004743>
12. Hansen, N.: The CMA evolution strategy: a comparing review. In: Lozano, J., et al. (eds.) *Towards a New Evolutionary Computation: Advances in Estimation of Distribution Algorithms*. Studies in Fuzziness and Soft Computing, vol. 192, pp. 75–102. Springer, Heidelberg (2006). [https://doi.org/10.1007/3-540-32494-1\\_4](https://doi.org/10.1007/3-540-32494-1_4)
13. Hansen, N., Auger, A., Finck, S., Ros, R.: Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical report, INRIA (2012). <http://coco.gforge.inria.fr/bbob2012-downloads>
14. Hansen, N.: Adaptive encoding: how to render search coordinate system invariant. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) *PPSN 2008*. LNCS, vol. 5199, pp. 205–214. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-87700-4\\_21](https://doi.org/10.1007/978-3-540-87700-4_21)
15. Hansen, N., Auger, A., Ros, R., Finck, S., Pošík, P.: Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In: *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation, GECCO 2010*, pp. 1689–1696. ACM, New York (2010). <http://doi.acm.org/10.1145/1830761.1830790>
16. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments - a survey. *IEEE Trans. Evol. Comput.* **9**(3), 303–317 (2005)
17. Kramer, O.: Evolution strategies with Ledoit-Wolf covariance matrix estimation. In: 2015 IEEE Congress on Evolutionary Computation (IEEE CEC) (2015)
18. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* **10**(5), 603–621 (2003)
19. Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. *J. Portf. Manage.* **30**(4), 110–119 (2004)

20. Ledoit, O., Wolf, M.: A well-conditioned estimator for large dimensional covariance matrices. *J. Multivar. Anal. Arch.* **88**(2), 265–411 (2004)
21. Ledoit, O., Wolf, M.: Non-linear shrinkage estimation of large dimensional covariance matrices. *Ann. Stat.* **40**(2), 1024–1060 (2012)
22. Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. Available at SSRN 2383361 (2014)
23. Meyer-Nieberg, S., Beyer, H.G.: On the analysis of self-adaptive recombination strategies: first results. In: McKay, B., et al. (eds.) *Proceedings of the 2005 Congress on Evolutionary Computation (CEC 2005)*, Edinburgh, UK, pp. 2341–2348. IEEE Press, Piscataway (2005)
24. Meyer-Nieberg, S., Beyer, H.G.: Self-adaptation in evolutionary algorithms. In: Lobo, F., Lima, C., Michalewicz, Z. (eds.) *Parameter Setting in Evolutionary Algorithms. Studies in Computational Intelligence*, vol. 54, pp. 47–76. Springer, Heidelberg (2007). [https://doi.org/10.1007/978-3-540-69432-8\\_3](https://doi.org/10.1007/978-3-540-69432-8_3)
25. Meyer-Nieberg, S., Kropat, E.: Small populations, high-dimensional spaces: Sparse covariance matrix adaptation. In: *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 525–535, September 2015
26. Meyer-Nieberg, S., Kropat, E.: Evolution strategies and covariance matrix adaptation - investigating new shrinkage techniques. In: *Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, pp. 105–116 (2016)
27. Meyer-Nieberg, S., Kropat, E.: Adapting the covariance in evolution strategies. In: *Proceedings of ICORES 2014*, pp. 89–99. SCITEPRESS (2014)
28. Meyer-Nieberg, S., Kropat, E.: A new look at the covariance matrix estimation in evolution strategies. In: Pinson, E., Valente, F., Vitoriano, B. (eds.) *ICORES 2014. CCIS*, vol. 509, pp. 157–172. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-17509-6\\_11](https://doi.org/10.1007/978-3-319-17509-6_11)
29. Meyer-Nieberg, S., Kropat, E.: Sparse covariance matrix adaptation techniques for evolution strategies. In: Bramer, M., Petridis, M. (eds.) *Research and Development in Intelligent Systems XXXII*, pp. 5–21. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-25032-8\\_1](https://doi.org/10.1007/978-3-319-25032-8_1)
30. Pourahmadi, M.: *High-Dimensional Covariance Estimation: With High-Dimensional Data*. Wiley, New York (2013)
31. Rechenberg, I.: *Evolutionstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag, Stuttgart (1973)
32. Rudolph, G.: Evolutionary strategies. In: Rozenberg, G., Bäck, T., Kok, J.N. (eds.) *Handbook of Natural Computing*, pp. 673–698. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-540-92910-9\\_22](https://doi.org/10.1007/978-3-540-92910-9_22)
33. Schwefel, H.P.: *Numerical Optimization of Computer Models*. Wiley, Chichester (1981)
34. Soysal, O., Sahin, E.: Probabilistic aggregation strategies in swarm robotic systems. In: *Proceedings 2005 IEEE Swarm Intelligence Symposium, SIS 2005*, pp. 325–332, June 2005
35. Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate distribution. In: *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, vol. 1, pp. 197–206 (1956)
36. Stein, C.: Estimation of a covariance matrix. In: *Rietz Lecture, 39th Annual Meeting. IMS, Atlanta, GA* (1975)
37. Thomaz, C.E., Gillies, D., Feitosa, R.: A new covariance estimate for bayesian classifiers in biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.* **14**(2), 214–223 (2004)

38. Tong, T., Wang, C., Wang, Y.: Estimation of variances and covariances for high-dimensional data: a selective review. *Wiley Interdisc. Rev. Comput. Stat.* **6**(4), 255–264 (2014)
39. Touloumis, A.: Nonparametric Stein-type shrinkage covariance matrix estimators in high-dimensional settings. *Comput. Stat. Data Anal.* **83**, 251–261 (2015)