UNIVERSITY OF DORTMUND

REIHE COMPUTATIONAL INTELLIGENCE

COLLABORATIVE RESEARCH CENTER 531

Design and Management of Complex Technical Processes and Systems by means of Computational Intelligence Methods

ICSPEA: Integrating CMA-ES and SPEA2

Jörn Mehnen, Tobias Wagner, Petra Kersting,
Igor Tipura and Günter Rudolph

No. CI-222/07

Technical Report ISSN 1433-3325 January 2007

Secretary of the SFB 531 · University of Dortmund · Dept. of Computer Science/LS 2
44221 Dortmund · Germany

This work is a product of the Collaborative Research Center 531, "Computational Intelligence," at the University of Dortmund and was printed with financial support of the Deutsche Forschungsgemeinschaft.

ICSPEA: Integrating CMA-ES and SPEA 2

Jörn Mehnen[1], Tobias Wagner[1], Petra Kersting[1], Igor Tipura[1], and Günter Rudolph[2]

[1]Institut für Spanende Fertigung (ISF)
[2]Chair of Algorithm Engineering
University of Dortmund, 44221 Dortmund, Germany
{mehnen, wagner, pkersting}@isf.de, guenter.rudolph@cs.uni-dortmund.de,
WWW home pages: http://www.isf.de, http://ls11-www.cs.uni-dortmund.de

Abstract. The covariance matrix adaptation (CMA) is a concept originally introduced for improving the single-objective evolution strategy (ES). CMA varies the classical ES-mutation operator by utilising a mutation distribution adaptation scheme and an evolution path, which takes the evolutionary history into account. SPEA 2 surely belongs to the most popular multi-objective evolutionary algorithms. It uses the strength Pareto concept and a special distribution measure for the evaluation of offspring individuals. An archive collects non-dominated individuals, which are used during selection, making the SPEA 2 a typical elitist strategy.
The new ICSPEA (Integrated CMA-SPEA) combines the powerful mutation concept of the CMA-ES with the evaluation scheme of the SPEA 2. Tests on selected benchmark functions show the promising features of this new multi-objective optimisation algorithm.

Keywords:
CMA (Covariance Matrix Adaptation), SPEA 2 (Strength Pareto Evolutionary Algorithm), Multi-objective Evolutionary Algorithms

1 Introduction

The evolution strategy (ES) is a flexible and powerful real value stochastic single-criterion optimisation strategy. The correlated mutations concept was introduced by Schwefel 1981 [1]. The idea is to adapt the n standard step sizes σ_i of the ES in order to raise the probability of producing successful mutation steps. Generally, the lines of equal probability density used in the ES form a n-dimensional hyper-ellipsoid. Without correlated mutations the main axes of this ellipsoid are aligned with the coordinate axes. The idea of correlated mutations is to turn this hyper-ellipsoid adaptively towards promising search directions using a $n \times n$ rotation matrix containing $(n^2 - n)/2$ rotation angles. This matrix undergoes variation and selection together with the step sizes and the objective variables assigned to each individual [2].

Hansen and Ostermeier identified several shortcomings of the implementation of the original correlated mutation concept [3]. They proposed a strongly

de-randomized single-criterion approach called Covariance Matrix Adaptation Evolution Strategy (CMA-ES). The algorithm utilises two information sources: a cumulation of selected mutation steps over time, the so called evolution path, and a covariance matrix, which is adapted by means of singular value decomposition of the evolution path. The CMA-ES is invariant against affine transformations such as rotation and translation of the search space [4]. Empirical tests show that on highly multimodal test functions, such as the scaled Rastrigin function, the CMA-ES performs better than the conventional ES with mutation control and ES with correlated mutations [3]. In the tests of Hansen and Ostermeier, the convergence speed and scaling of the CMA-ES was also excellent, i.e. it scales between linear to quadratic with the dimensions of the problems. Especially on non-separable test functions the CMA-ES should be favoured over the ES with global or individuals step size adaptation, only [3].

Multi-objective problems can be approached with single-objective techniques. These approaches are useful if enough a-priori knowledge is given or if there is no demand for compromising solutions (see [5] or [6]). If the user is more interested in alternative solutions, it is more appropriate to use multi-objective strategies [7], [8]. Multi-objective evolutionary algorithms (MOEA) are able to produce a complete Pareto front in one single run. A technique to utilize user preferences is to guide the MOEA with desirability functions (see [9], [10], and [11]).

There exist a huge number of different multi-objective evolutionary algorithms by now. The best known algorithms are NSGA-II (Non-dominated Sorting Genetic Algorithm, Version 2) introduced by Deb et al. [12] and SPEA 2 (Strength Pareto Evolutionary Algorithm, Version 2) by Zitzler, Laumanns, and Thiele [13]. These algorithms are widely accepted and often taken as reference for comparison. NSGA-II and SPEA 2 mainly differ in their fitness assignment and the selection scheme. SPEA 2 uses an additional archive for storing non-dominated individuals while NSGA-II does not use an archive at all. SPEA 2 and NSGA-II are both basically elitist strategies, i.e. both algorithms evaluate offspring together with parents or the archive, respectively. In the original SPEA 2 the individuals were coded as real vectors using polynomial mutation and the SBX-20 recombination operator [14]. The results show that SPEA 2 and NSGA-II perform similarly well on typical test functions. SPEA 2 provides a better point distributions while NSGA-II reaches a broader spread of points along the Pareto front. NSGAII is faster than SPEA 2 with respect to worst-case complexity. In higher dimensional objective spaces SPEA 2 seems to be advantageous compared to NSGA-II [13], what makes it highly relevant for practical applications.

Igel, Hansen, and Roth propose the multi-objective MO-CMA-ES [4], which is an implementation of λ_{MO} CMA-ES working in parallel. Each CMA-ES generate λ offspring from a pool of λ_{MO} parents using the CMA mutation scheme. For selection, parents together with offspring are ordered using either a crowding distance (as used in the NSGA-II) – referred as c-MO-CMA – or a contributing hyper-volume relation – referred as s-MO-CMA [4]. The best of offspring and parent individuals form the next generation of λ_{MO} individuals. This makes the

MO-CMA-ES an elitist strategy. The covariance matrices are updated in parallel and independently from each other within each single-objective $(1+\lambda)$-CMA-ES. In the work of Igel, Hansen, and Roth the MO-CMA-ES has been compared with the NSGA-II, which uses a polynomial mutation and the SBX-20 recombination operator. The NSDE (non-dominated sorting differential evolution) algorithm of Iorio and Li [15] also competed in the comparison. A hyper-volume indicator [16] and an epsilon indicator [17] have been used to compare the Pareto fronts. The results show that especially the s-MO-CMA generally performs statistically significantly better than NSGA-II on Fonseca's function, ZDT-functions [18] and spherical functions with respect to both performance measures. Only on ZDT4 the NSGA-II performs better because this function fits perfectly to the NSGA-II variation operators. The NSDE did not perform very well compared to the other algorithms. c-MO-CMA works well an Fonseca's function and the spherical models. The unfavourable performance of NSGA-II and c-MO-CMA may be due to the fact that these algorithms use the crowding distance operator in their evaluation scheme. This operator may be good to increase the distribution of the solutions but not favourable for progress in terms of selecting better solutions [4].

An analysis of the CMA and MO-CMA shows that the CMA-technique has advantageous over the conventional ES and NSGA-II, respectively. A drawback in NSGA-II with respect to s-MO-CMA seems to be the evaluation operator. SPEA 2 has advantageous properties with respect to higher objective space dimensions and point distributions along the Pareto front. Laumanns, Rudolph, and Schwefel [19] showed on the basis of an MO-ES with dominance evaluation and a multi-objective predator-prey-model that proper adaptation strategies of the step sizes are especially important for multi-objective problems approaching the Pareto front.

Looking at the different results of the analyses discussed before, it seems to be necessary to combine proper variation and evaluation operators to form a new multi-objective optimisation algorithm. CMA has a very favourable variation scheme while SPEA 2 shows superior results especially due to its evaluation scheme. In the following the respective operators of CMA and SPEA 2 have been joined in order to form the new ICSPEA algorithm. In the following, this new method will be introduced in more detail. A brief empirical comparison with SPEA 2 on standard benchmark problems is given to get a first impression of the algorithm's features.

2 Concept

The real-value multi-objective optimisation algorithm ICSPEA consists of a CMA-core, which is used for the variation of the individuals, and a SPEA 2-core for the evaluation of the individuals. The truncation selection scheme known from the evolution strategy realises a flexible method for controlling the selection pressure. An archive is used to store non-dominated individuals.

Encoding The problems, which can be approached by ICSPEA, are arbitrary mappings from the n-dimensional real-value decision space subset $X \subseteq \mathbb{R}^n$ into the m-dimensional real-value objective space $Y \subseteq \mathbb{R}^m$. The m-criterion problem can be stated as:

$$\begin{array}{l} 'min' \ (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^T \quad \text{with} \ \ \mathbf{x} \in \mathbb{R}^n. \\ \mathbf{x} \in \mathbf{X} \end{array} \tag{1}$$

Each individual is a sample from the decision space $\mathbf{x} \in \mathbb{R}^n$. In the following, minimisation of each objective is the general goal. The ICSPEA is an a-posteriori method that generates a discrete approximation of a Pareto front rather than finding a single minimum solution. References on the concept of Pareto domination can be found in [6] or [5]. A good distribution of the discrete approximations of the Pareto front in combination with its fast and exact approximation independently from the size of the decision and objective space is the general goal of a-posteriori multi-objective evolutionary algorithms.

Mutation and Recombination ICSPEA utilises the variation operators of the CMA-ES. New individuals $\mathbf{x}_{i(g+1)}, i = 1, \ldots, \lambda, \lambda \in \mathbb{R}_+$ are sampled using a n-dimensional normal distribution [20].

$$\mathbf{x}_{i(g+1)} \propto N_i(\mathbf{m}_{(g)}, \sigma^2_{(g)}\mathbf{C}_{(g)}) = \mathbf{m}_{(g)} + \sigma_{(g)} N_i(\mathbf{0}, \mathbf{C}_{(g)}), \tag{2}$$

where (g) is the generation counter, $\sigma_{(g)} \in \mathbb{R}_+$ is called step size, and $\mathbf{m}_{(g)} \in \mathbb{R}^n$ is the mean of the distribution. $\mathbf{C}_{(g)} \in \mathbb{R}^{n \times n}$ determines the orientation of the distribution ellipsoid.

The distribution mean $\mathbf{m}$ of each individual $\mathbf{x}_i = \mathbf{m} + \sigma N_i(\mathbf{0}, \mathbf{C})$ is updated by a ranked selection of the μ 'best' individuals and the application of a weighted intermediate recombination. The individuals are ordered according to their strength Pareto values (see below). The μ best individuals x'_i form a new set D. Depending on the strategy, the individuals are chosen from the set of λ offspring only (non-elitist) or from the united sets of offspring and archive (SPEA 2 like elitist). An archive of size α stores (mainly) the non-dominated individuals (see below). The update strategy of the new mean m of each individual from generation (g) to the next generation $(g+1)$ is:

$$\mathbf{m}_{(g+1)} = \sum_{i=1}^{\mu} w_i \mathbf{x'}_{i(g+1)} = \mathbf{m}_{(g)} + \sigma_{(g)} \sum_{i=1}^{\mu} w_i N_i(\mathbf{0}, \mathbf{C}_{(g)}) \tag{3}$$

The μ weights $w_i \in \mathbb{R}_+$ can be arbitrarily chosen but must sum up to one.

The covariance matrix $\mathbf{C}$ is updated from generation g to $g+1$ using the so called evolution path [3]. Actually, a sequential principle component analysis is conducted. On quadratic functions, $\mathbf{C}$ approximates the inverse Hessian matrix. The idea is to adapt $\mathbf{C}$ in a way that new successful steps are more likely to appear during the following generations. The distribution of $\mathbf{C}$ is updated by accounting for already successful steps found. The influence of each successful step

decreases exponentially. The cumulated path $\mathbf{p}_c$ of successful steps is updated according to the formula [21]:

$$\mathbf{p}_{c(g+1)} = (1 - c_c)\ \mathbf{p}_{c(g)} + h_{\sigma(g+1)} \sqrt{c_c(2 - c_c)\mu_{\text{eff}}}\ \frac{\mathbf{m_{(g+1)}} - \mathbf{m_{(g)}}}{\sigma_{(g)}}, \tag{4}$$

where $c_c << 1$ and $\mu_{\text{eff}} = \sum_{i=1}^{\mu} w_i^2$ with $\sum_{i=1}^{\mu} w_i = 1$. The value μ_{eff} is termed variance effective selection mass [20]. The parameter $h_{\sigma(g+1)}$ is used to stop the increase of $\mathbf{p}_{\sigma(g+1)}$, if the length of the evolution path $||\mathbf{p}_{\sigma(g+1)}||$ is too large. The Heaviside function is defined following [21]:

$$h_{\sigma(g+1)} := \begin{cases} 1 & \text{if} \quad \frac{||\mathbf{p}_{\sigma(g+1)}||}{\sqrt{1-(1-c_\sigma)^{2(g+1)}}} < (1.5 + \frac{1}{n-0.5} E[||N(\mathbf{0}, \mathbf{I})||]) \\ 0 & \text{if} \quad \text{otherwise} \end{cases} \tag{5}$$

A rank-μ updating of the correlation matrix $\mathbf{C}_{(g+1)}$ is realized by the weighted outer product of the μ best selected steps $\mathbf{x'}_i$. The adaptation of $\mathbf{C}_{(g+1)}$ follows the equation

$$\begin{aligned} \mathbf{C}_{(g+1)} = (1 - c_{cov})\ \mathbf{C}_{(g)} + \frac{c_{cov}}{\mu_{cov}} (\mathbf{p}_{c(g+1)} \mathbf{p}_{c(g+1)}^T + [(1 - h_{\sigma(g+1)}) c_c(2 - c_c)] \mathbf{C}_{(g)}) \\ + c_{cov}(1 - \mu_{cov}^{-1})\ \mathbf{Z}, \end{aligned} \tag{6}$$

where

$$\mathbf{Z} = \sum_{i=1}^{\mu} w_i\ \frac{1}{\sigma_{(g)}} [(\mathbf{x'}_{i(g+1)} - \mathbf{m}_{(g)})(\mathbf{x'}_{i(g+1)} - \mathbf{m}_{(g)})^T]. \tag{7}$$

The path length is adapted according to the next formula:

$$\mathbf{p}_{\sigma(g+1)} = (1 - c_\sigma)\ \mathbf{p}_{\sigma(g)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{\text{eff}}}\ \mathbf{C}_{(g)}^{-\frac{1}{2}} \frac{\mathbf{m_{(g+1)}} - \mathbf{m_{(g)}}}{\sigma_{(g)}} \tag{8}$$

The step sizes are adapted following the path length control scheme [3]:

$$\sigma_{(g+1)} = \sigma_{(g)}\ exp\left(\frac{c_\sigma}{d_\sigma}\left(\frac{||\mathbf{p}_{\sigma(g+1)}||}{E[||N(\mathbf{0}, \mathbf{I})||]} - 1\right)\right) \tag{9}$$

A good approximation for $E[||N(\mathbf{0}, \mathbf{I})||] = \sqrt{2}\,\Gamma(\frac{n+1}{2})/\Gamma(\frac{n}{2})$ is $E[||N(\mathbf{0}, \mathbf{I})||] \approx \sqrt{n}(1 - \frac{1}{4n} - \frac{1}{21\,n^2})$ [3].

The values c_c and c_{cov} are learning rates. The larger the values the faster the learning and the faster the older data is disregarded. Initially, the covariance matrix is set to the unity matrix $\mathbf{I} \in \mathbb{R}^{n \times n}$ and $\mathbf{p_{c}}_{(g+1)} = \mathbf{0} \in \mathbb{R}^{\mathbf{n}}$. The rank-$\mu$ updating is especially useful in large populations with $\lambda > 3n + 10$ and $d_\sigma \approx 1 + \sqrt{\frac{\mu_{\text{eff}}}{n}}$ [20]. Hansen recommends $c_c \approx 4/n$, $c_\sigma \approx 4/n$, $c_{cov} = \frac{\mu_{\text{eff}}}{n^2}$, and $\mu_{cov} = \mu_{\text{eff}}$ [20]. The recommended number of parents is $2 \leq \mu \leq n$ and the number of offspring is $0.27\mu \leq \lambda \leq 0.5\mu$ [3]. The new offspring population is termed $P'_{(g+1)}$.

Evaluation After recombination and mutation of the individuals the multi-objective evaluation scheme of the SPEA 2 assigns a scalar measure, the strength $S(\mathbf{x}_i)$, to each individual. The individuals ($\mathbf{x}_i$) selected for the evaluation depends on the strategy of ICSPEA. Generally, the selection scheme can use the offspring population as well as the archive and the parent population in different combinations. The strengths are calculated as follows:

$$S(\mathbf{x}_i) = |\{\mathbf{x}_j | \mathbf{x}_j \in P'_{(g+1)} \wedge \mathbf{x}_i \prec \mathbf{x}_j\}|, \tag{10}$$

i.e. counting the individuals j that are dominated by the individual $\mathbf{x}_i$. Other alternatives could be to evaluate archive and offspring, parents and offspring or all three sets of individuals together. In the current ICSPEA implementation the strength is evaluated on the offspring population $P'_{(g+1)}$ only. Analyses of other combinations are matter of current research. The raw fitness $R(\mathbf{x}_i)$ of individual $\mathbf{x}_i$ is the sum of strengths of all individuals $\mathbf{x}_j$ the individual $\mathbf{x}_i$ is dominating.

$$R(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in P'_{(g+1)}, \mathbf{x}_j \prec \mathbf{x}_i} S(\mathbf{x}_j) \tag{11}$$

The archive $\bar{P}_{(g)}$ is initially assigned the empty set. For later generations it is updated as described below.

Additionally to the raw fitness, a density function is introduced which helps to increase the distribution spread of the solutions along the approximated Pareto front. The density function is the k-th nearest neighbourhood method as used by SPEA 2. All distances of each individual i to its neighbours in $P'_{(g+1)}$ are assigned the respective individual and the list is sorted in increasing order. The k-th distance in the list of individual $\mathbf{x}_i$ is sought for, termed σ_i^k. The authors [13] recommend $k = \sqrt{N + \bar{N}}$ in their SPEA 2 algorithm, where N is the size of the offspring population and $\bar{N}$ the current size of the archive. For the comma-strategy of course $k = \sqrt{N}$ has to be used.

The density function in SPEA 2 corresponding to $\mathbf{x}_i$ is defined as follows:

$$D(\mathbf{x}_i) = \frac{1}{\sigma_i^k + 2} \tag{12}$$

In the ICSPEA a more expensive alternative has been applied. This version puts a stronger emphasize on the relative distance between each of the individuals.

$$D'(\mathbf{x}_i) = \sum_{j=1}^{k} \frac{1}{\sigma_i^j + 2^j} \tag{13}$$

The total fitness $F(\mathbf{x}_i) : \mathbb{R}^n \to \mathbb{R}$ of each individual $\mathbf{x}_i$ is

$$F(\mathbf{x}_i) = R(\mathbf{x}_i) + D'(\mathbf{x}_i). \tag{14}$$

Archiving and Selection ICSPEA has an archive $\bar{P}_{(g)}$ of fixed size α. After selection, the archive is updated. In the first generation the archive is initialised the empty set. After the evaluation step, the individuals from the offspring population and the individuals from the current archive $\bar{P}_{(g)}$ are united and the non-dominated individuals are selected from the united set. The old archive is deleted and the new $\bar{P}_{(g+1)}$ is filled with non-dominated individuals. If the number of non-dominated individuals is exactly the size of the archive, the archiving is finished. If the number of non-dominated individuals is smaller than the archive size α, the remaining space is filled with individuals with slightly smaller fitnesses. If the number of non-dominated individuals is larger than the archive size, then an archive truncation procedure is invoked [13], which chooses individuals with maximum distance to others in the archive. This ensures a good distribution of solutions within the archive.

The truncation selection scheme of the classical evolution strategy has been applied for the ICSPEA. In order to find the most promising offspring population, the individuals' fitness values are evaluated and sorted. The μ best with respect to their total scalar fitness F are passed on to the next generation. It should be pointed out that the total fitness F assigned to each solution by the SPEA 2 concept is actually a scalar value although the problem itself is multi-objective. It depends on the strategy if individuals from the archive together with the offspring population are inspected or just the offspring population is taken into account. The new μ individuals selected replace the former parent population to become $P_{(g+1)}$.

Termination Criterion The cycle of mutation and recombination, evaluation, archiving, and selection is repeated until a stopping criterion is satisfied. Like SPEA 2, ICSPEA uses a limited number of generations. The result of the optimization procedure is the set A of solutions in decision space that corresponds to the non-dominated individuals in the archive $\bar{P}_{(g)}$. In order to visualize the Pareto front the individuals with their corresponding m fitness values are displayed.

3 Experimental Setup

ICSPEA is tested on several typical benchmark problems. The functions have been selected to show the special features of the algorithm. Schaffer's function and a variation with variable dimension decision space are analysed to reveal features that are typical for spherical functions. The Pareto front of the complex Kursawe problem has a convex-concave shape and is partially non-connected.

Each experiment has been carried out 50 times starting with different initial positions and different random number generator initialisations. Box-and-whiskers plots show the basic statistical behaviour of the solutions. A statistical parameter optimisation of the algorithm has not been performed. In order to be able to estimate the quality of the algorithm, a standard SPEA 2 from the toolbox PISA [22] has been used. For the SPEA 2, tests with empirically best known parameter settings have been applied.

3.1 Test Functions

In the results, first, ICSPEA is applied to **Schaffer's function** [23]. This function has a one-dimensional decision space with two dimensions in the objective space. The Pareto set is the compact interval $[0, 2]$. The function has been selected due to its basically quadratic features.

$$\text{SCH}_1: \qquad \begin{array}{l} f_1(x) = x^2 \\ f_2(x) = (x-2)^2 \end{array} \tag{15}$$

with restriction $x \in [-10^6, 10^6]$.

A **generalised Schaffer's function** for higher dimensions $n > 1$ of the decision space is used. The original version with $n = 2$ was introduced by [24]. The variant SCH_n for variable dimensions n has the Pareto set $x_1 \in [0, 2]$ and $x_2 = \ldots = x_n = 0$.

$$\text{SCH}_n: \qquad \begin{array}{l} f_1(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \\ f_2(\mathbf{x}) = (x_1-2)^2 + \sum_{i=2}^{n} x_i^2 \end{array} \tag{16}$$

with restriction $x_1, x_2 \in [-10^6, 10^6]^2$.

The two-criterion **Kursawe function** [25] has a mixed-shaped convex-concave Pareto front. In the tests with ICSPEA the number of dimensions have been set to $n = 20$. The area of feasible solutions is $[-100.0, 100.0]^n$.

$$f_1(\mathbf{x}) = \sum_{i=1}^{n-1} -10e^{(-0.2 \cdot \sqrt{x_i^2 + x_{i+1}^2})} \tag{17}$$

$$f_2(\mathbf{x}) = \sum_{i=1}^{n} (|x_i|^{0.8} + 5\sin(x_i)^3) \tag{18}$$

Convergence assessment Due to the fact that the Pareto sets of both functions SCH_1 and SCH_n are known, the average distance of all current solutions in the archive of size α to the line $[0, 2]$ can be taken for estimating the convergence speed. Like in single-objective optimisation the distance of the solutions to known best solutions is calculated. Here the measures MeSCH_1 and MeSCH_n, respectively, determine the average distance of the approximating Pareto set to the true Pareto set.

In the case of SCH_1, the distance measure for each one-dimensional individual x_i, $i = 1, \ldots, \alpha$ is:

$$d(x_i) = \begin{cases} (x_i - 2) & \text{if} \quad (x_i > 2) \\ -x_i & \text{if} \quad (x_i < 0) \\ 0.0 & \text{if} \quad (0 \le x_i \le 2) \end{cases} \tag{19}$$

$$\text{MeSCH}_1 = \tfrac{1}{\alpha} \sum_{i=1}^{\alpha} d(\{x_i\})$$

The measure MeSCH_n works for experiments with SCH_n. Here the distance of n-dimensional individuals $\mathbf{x}_i$, $i = 1, \ldots, \alpha$ to the line structure $[0, 2] \times \mathbf{0}^{n-1}$ has to be measured. The first component $\mathbf{x}_{i1}$ of each individual $\mathbf{x}_i$ is treated as in MeSCH_1 yielding $d(\mathbf{x_i})$. The remaining components are projected to the coordinate axes and the total of the absolute lengths of these projections is added to $d(\mathbf{x_i})$. Like in MeSCH_1, the average of the total of all distances is measured for the complete archive.

$$d(\mathbf{x_i}) = \begin{cases} (\mathbf{x_{i1}} - 2) & \text{if} \quad (\mathbf{x_{i1}} > 2) \\ -\mathbf{x_{i1}} & \text{if} \quad (\mathbf{x_{i1}} < 0) \\ 0.0 & \text{if} \quad (0 \leq \mathbf{x_{i1}} \leq 2) \end{cases} \tag{20}$$

$$\text{MeSCH}_n = \tfrac{1}{\alpha} \sum_{i=1}^{\alpha} \sqrt{d(\mathbf{x_i})^2 + \sum_{j=2}^{n} \mathbf{x_i}_j^2}$$

Convergence measures for Kursawe's functions have not been done. References on quality assignments for Pareto fronts can be found in [26], [8], or [27].

3.2 Parameter Settings

Table 1 shows the parameter settings used. $\mu \in \mathbb{N}_+$ defines the size of the panmictic parent population, $\lambda \in \mathbb{N}_+$ is the size of the offspring. In case of a comma-strategy, the value of λ must be bigger than the number μ of parents. The experiments feature only the comma strategy. $\alpha \in \mathbb{N}_+$ is the size of the archive. $\sigma_0 \mathbb{R}_+$ is the initial step size of the algorithm. The internal covariance matrix is initialized $\mathbf{C}_{(0)} = \mathbf{I}$ and the evolutionary paths $\mathbf{p}_{c\,(0)} = \mathbf{0}$ and $\mathbf{p}_{\sigma\,(0)} = \mathbf{0}$, respectively. As recommended by [3] the initial distribution mean $\mathbf{m}_{(0)}$ should be set generally problem depending within the area of feasible solutions $[min, max]^n$. The value $k \in \mathbb{N}_+$ is used to determine the distance measure $D(\mathbf{x}_i)$ utilized in the strength Pareto evaluation. The maximum number of generations is $gen \in \mathbb{N}_+$. This value also depends on the problem. In PISA the SPEA 2 algorithm uses a fixed $k = 1$. In the experiments the specific SPEA 2 parameters $p_m = \frac{1}{n}$, $p_c = 0.5$, $\eta_m = 20$, and $\eta_m = 15$ were used and the SBX-variation operator was applied.

Table 1: The ICSPEA parameter settings

Problem	μ	λ	α	σ_0	$\mathbf{m}_{(0)}$	k	Gen.	dim.
SCH_1	5	50	60	375.0	uniform in $[-10^6, 10^6]$	10	40	1
SCH_n	5	50	60	375.0	uniform in $[-10^6, 10^6]^n$	10	400	20
Kursawe	5	50	60	1.0	uniform in $[-100.0, 100.0]^n$	10	40	2
Kursawe	5	50	60	1.0	uniform in $[-100.0, 100.0]^n$	10	40	10

4 Results

ICSPEA has been applied to Schaffer's function with one-dimensional and 20-dimensional decision space. The convergence speed of the algorithm is nearly linearly in logarithmic scaling during the first 30 steps (see Figure 1 on the left). After reaching the Pareto set only slight variations appear.

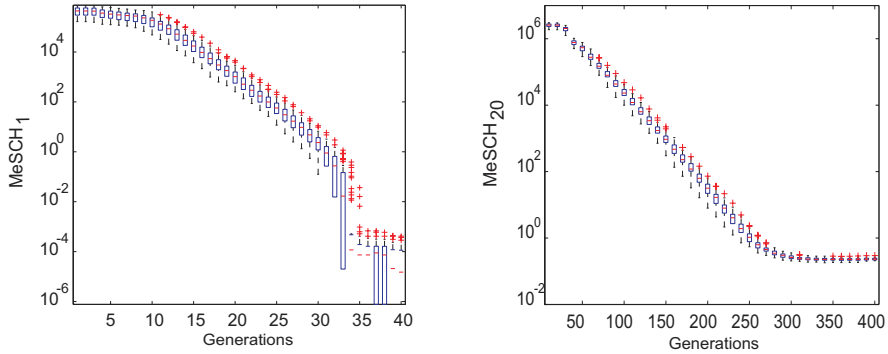

Figure 1: Convergence plots for ICSPEA with 5 parents and 50 offspring applied to the one-dimensional (left) and 20-dimensional Schaffer function (right). An archive of size 60 was used.

The linear convergence supports the hypothesis that the CMA step size adaptation performs very well on sphere like problems. Although the fitness values assigned to the individuals during the SPEA 2 evaluation are not absolutely quadratic, the basically quadratic characteristics of the partial functions f_1 and f_2 of Schaffer's problem may be the reason for the good performance of ICSPEA in this simple problem. The small heights of the boxes in the figures indicate that the runs show only slight variations. Large outliers do not appear in all 50 runs. In some cases the algorithms found the true Pareto set (e.g. after about 37 generations), i.e. all values of the individuals are in the interval $[0, 2]$.

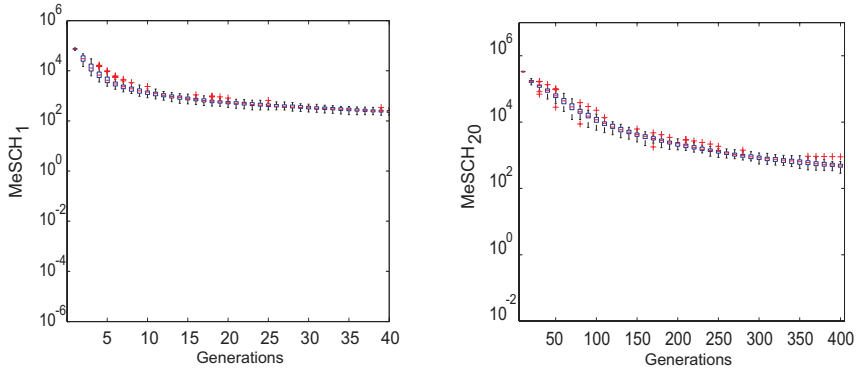

Figure 2: Convergence plots for $(5 + 50)$-SPEA 2 for the one-dimensional Schaffer function after 40 generations (left). The same algorithm applied to 20-dimensional Schaffer function after 400 generations (right).

The 20-dimensional Schaffer problem is more difficult, because all values in the genome have to be zero except the first parameter, which has to be in the $[0, 2]$ interval. In figure 1 on the right the values for MeSCH_1 have been plotted with logarithmic scaling. A box plot for every tenth generation is shown. The median values after generation 400 do not change any more for higher generation numbers. They have been omitted in the plot. The results show that the variations over all 50 runs are quite small. Small deteriorations of the fitness values of MeSCH_1 are due to the characteristics of the SPEA evaluator that intends to spread the solutions along the Pareto front. Some solutions, which are already within the desired solution area, are 'pushed out' of the interval again yielding a slight deterioration in the fitness values.

It is interesting to compare ICSPEA with results of the standard SPEA 2 from the software package PISA [22]. For the evaluation of the approximation quality of SPEA 2 again the measures MeSCH_1 and MeSCH_{20} have been applied. One can see in figure 2 that the step size adaptation of SPEA 2 leads to a very early stagnation of the convergence towards the Pareto front. As already mentioned in [19] the SBX technique is not optimal. An alternative to the technique proposed by [19] is CMA. The application of the CMA in ICSPEA increases the power of the standard SPEA 2 as can be seen in figure 1. The advantageous features coming from the strength Pareto selection such as the even distributions of the solutions along the Pareto front is maintained in ICSPEA. This can be seen in the following figures.

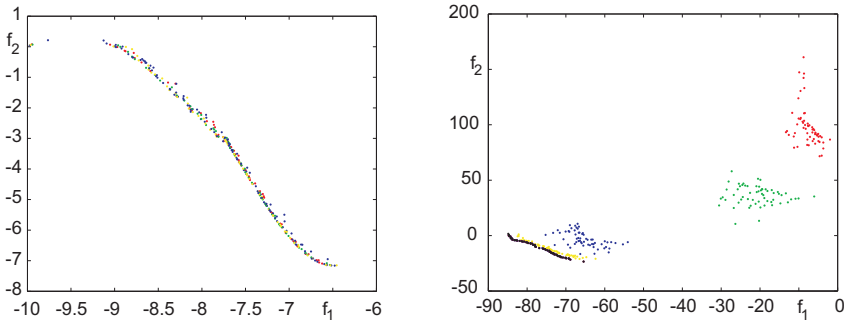

Figure 3: Five different runs approximating the Pareto front of Kursawe's function with a two dimensional decision space using a ICSPEA with 5 parents and 50 offspring (left). Approximation of the Pareto front of Kursawe's function with a 10-dimensional decision space. The generations 40 (red), 60 (green), 80 (blue), 100 (yellow) to 800 (black) are displayed.

The function of Kursawe is more complex than Schaffer's function. The functions f_1 and f_2 are nonlinear and the Pareto front has a partially convex and concave shape. The complexity of the problem was varied by changing the dimension of the decision space from 2 to 10. Figure 3 shows that in both dimensions ICSPEA is able to approximate the Pareto front nicely. The shape and values are identical to [25]. The SPEA 2 evaluation scheme is, as expected, not influenced by the shape of the Pareto front. The scheme finds equally distributed

solutions along the Pareto front. In figure 3 various Pareto front approximations are shown. ICSPEA generates approximations with only few variations of the results. The convergence of the populations towards the Pareto front in the case of a 10-dimensional decision space is shown in figure 3 on the right.

4.1 Summary

The research on the Covariance Matrix Adaptation (CMA) in single-objective evolution strategies yielded very promising variation operators with a nice step size adaptation mechanism. The SPEA 2 algorithm is supplied with a powerful evaluation method that helps to find well spread Pareto front approximations. First tests on typical benchmark functions show that the combination of both concepts seems to be a very promising concept. A proper parameter adaptation scheme is surely a key issue in optimisation. In ICSPEA, especially the CMA step size adaptation mechanism is emphasised. The key idea in ICSPEA is – as long as no a-priori knowledge is available – that all data gathered during the stochastic search process should be used as efficiently as possible to increase the success probability of a multi-objective optimization algorithm.
In order to further improve ICSPEA, a sound statistical analysis of the best initial parameter settings should be performed. Alternative methods for the evaluation of parents, offspring, and archive are matter of current research.

Acknowledgements

This work was supported by the *Deutsche Forschungsgemeinschaft (DFG)* as part of the *Collaborative Research Center 'Computational Intelligence' (SFB 531).*

References

1. Schwefel, H.P.: Numerical Optimization of Computer Models. Wiley, Chichester, England (1981)
2. Schwefel, H.P.: Evolution and Optimum Seeking. John Wiley, New York (1995)
3. Hansen, N., Ostermeier, A.: Completely Derandomized Self-Adaptation in Evolution Strategies. Evolutionary Computation **9**(2) (2001) 159–195
4. Igel, C., Hansen, N., Roth, S.: The Mutli-Objective Variable Metric Evolution Strategy, Part I. Technical Report IR-INI 2005-04, Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany (2005)
5. Miettinen, K.: Nonlinear Multiobjective Optimization. Int. series in operations research and management science. Kluwer Academic Publishers, Boston, USA (1998)
6. Ehrgott, M.: Multicriteria Optimization. Volume 491 of Springer, Lecture Notes in Economics and Mathematical Systems. Springer, Berlin (2000)
7. Deb, K.: Multi-Objective Optimization using Evolutionary Algorithms. Wiley, New York (2001)
8. Collette, Y., Siarry, P.: Multiobjective Optimization. Principles and Case Studies. Cost Engineering in Practice. Springer, Berlin (2003)

9. Branke, J., Deb, K.: Integrating user preferences into evolutionary multi-objective optimization. In Jin, Y., ed.: Knowledge Incorporation in Evolutionary Computation, Berlin, Springer (2004) 461–478
10. Trautmann, H., Mehnen, J.: A Method for Including A-Priori-Preferences in Multicriteria Optimization. Technical Report 49/2005, University of Dortmund, Germany, SFB 475 (2005)
11. Mehnen, J., Trautmann, K.: Integration of Expert's Preferences in Pareto Optimization by Desirability Function Techniques. In Teti, R., ed.: Proceedings of the 5th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering, ICME 2005, Ischia, Italy, University of Naples, Italy (2004) 293–298
12. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multi-Objective Genetic Algorithm-NSGA-II. Technical Report 2000001, KanGAL (2000)
13. Zitzler, E., Laumanns, M., Thiele, L.: SPEA 2: Improving the Strength Pareto Evolutionary Algorithm. Technical Report 103, Computer Engineering and Communication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Gloriastrasse 35, CH-8092 Zürich (2001)
14. Deb, K., Agarwal, R.: Simulated Binary Crossover for Continuous Search Space. Complex Systems **9** (1995) 115–148
15. Ioro, A., Li, X.: Solving Rotated Multiobjective Optimization Problems Using Differential Evolution. In Webb, Yu, eds.: Proceedings of the 17th Joint Australian Conference on Artificial Intelligence, LNCS 3339, Springer Verlag (2005) 861–872
16. Zitzler, E., Thiele, L.: Multiobjective Optimization using Evolutionary Algorithms – A Comparative Case Study. In Eigen, Bäck, Schoenauer, Schwefel, eds.: Fifth International Conference on Parallel Problem Solving From Nature (PPSN-V), Springer Verlag (1998) 292–301
17. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, M., de Fonseca, V.G.: Performance Assignment of Multiobjective Optimizers: An Analysis and Review. IEEE Transactions on Evolutionary Computation **7**(2) (2003) 117–132
18. Zitzler, E., Deb, K., Thiele, L.: Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. Evolutionary Computation **8**(2) (2000) 173–195
19. Laumanns, M., Rudolph, G., Schwefel, H.P.: Mutation Control and Convergence in Evolutionary Multi-Objective Optimization. In Matousek, R., Osmera, P., eds.: Proceedings of the 7th International Mendel Conference On Soft Computing (MENDEL 2001), Brno, Czech Republic, University of Technology (2001) 24–29
20. Hansen, N.: Tutorial: Covariance Matrix Adaptation (CMA). Tutorial, ETH-Zürich, Held at PPSN 2006 (2006)
21. Hansen, N.: The CMA Evolution Strategy: A Tutorial. Technical report, ETH-Zürich (2005)
22. Bleuler, S., Laumanns, M., Thiele, L., Zitzler, E.: PISA — A Platform and programming language independent Interface for Search Algorithms. In: EMO2003. (2003) 494–508
23. Schaffer, J.: Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. In: Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms, Lawrence Erlbaum (1985) 93–100
24. Laumanns, M., Rudolph, G., Schwefel, H.P.: A Spatial Predator-Prey Approach to Multi-Objective Optimization: A Preliminary Study. In Bäck, T., Eiben, A., Schoenauer, M., Schwefel, H.P., eds.: Parallel Problem Solving From Nature - PPSN V, Berlin, Springer (1998) 241–249
25. Kursawe, F.: A Variant of Evolution Strategies for Vector Optimization. In Schwefel, H.P., Männer , R., eds.: Parallel Problem Solving from Nature. 1st Workshop,

PPSN I. Volume 496 of Lecture Notes in Computer Science., Berlin, Germany, Springer-Verlag (1991) 193–197
26. Coello Coello, C., van Veldhuizen, D., Lamont, G.: Evolutionary Algorithms for Solving Multi-Objective Problems. Kluwer Academic Publishers, New York (2002)
27. Mehnen, J.: Mehrkriterielle Optimierverfahren für produktionstechnische Prozesse. Vulkan Verlag, Essen, Germany (2005) Habilitationsschrift.