Evolution Strategies with Line Search for Structural Optimization

Niels Bohn, Ulrich Gabbert

Niels Bohn, Ulrich Gabbert, Institute of Mechanics, Chair of Computational Mechanics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany
niels.bohn@bmu.de, ulrich.gabbert@mb.uni-magdeburg.de

1. Abstract
Optimization tasks in structural mechanics are often described by non-smooth, non-differentiable and multimodal objective functions. Evolution strategies (ES) are an appropriate choice for solving such problems, since the method is robust, general applicable and well suited for global optimization. The disadvantage of evolution strategies lies in the high number of objective function evaluations being necessary to locate the optimum. For applications in structural mechanics, a new variant of evolution strategies, the evolution strategies with line-search (L-ES), is developed. The method exhibits the mentioned good properties of the standard ES and reduces for certain problem classes the number of needed objective function evaluations. The L-ES is based on the standard ES. In some iteration steps, approximations of descent directions are computed by using information from successful past iteration steps. Extrapolations are performed along these directions, where the extrapolation step size is determined by a new self-adaptive scheme which requires no additional objective function calls. The L-ES can be used with the full range of possible parameterizations of the standard evolution strategies, and is not restricted to special types of optimization problems. Mathematical benchmark functions are employed to evaluate the characteristics of the new method and an application from structural mechanics is presented.

2. Keywords: evolution strategies, line search, gradient approximation, self adaptation

3. Introduction
Evolution strategies can be classified as a type of direct methods, requiring only information about the objective function during the search for the optimum. Contrary to analytical methods, the calculation of gradients is not necessary. The algorithm is robust, i.e., it can be applied to a wide range of problem classes successfully, has good global search properties and is able to cope with non-smooth and noisy objective functions. Such methods are well suited for practical engineering applications, where in the most cases the properties of the objective functions are unknown. The disadvantage of the ES lies in the high number of objective function evaluations which is necessary to locate the optimum. If the objective function has to be evaluated by a finite element analysis, the computational cost for the optimization can be enormously high. Thus various methods to improve the performance of the ES have been developed [1, 2, 3, 4, 5, 6, 7]. Adaptive approximation models for the objective function where implemented by [8, 9, 10], and hybrid methods, coupling the ES with other optimization algorithms were suggested by [11, 8, 12, 13]. In this paper we propose a new variant of the ES termed evolution strategies with line search. The aim of the investigation is a decrease of the computational cost at least for certain problem classes without losing, in general, the algorithm’s good global search properties and its robustness. The modifications are implemented in such a way, that the L-ES is not restricted a-priori to special problem classes or parameterizations of the evolution strategies. Thus, the full capabilities of the standard ES can be exploited. The paper is organized as follows. In section 4 the evolution strategies with line search are introduced and compared with the standard ES on the basis of a number of mathematical test functions. An engineering application is presented in section 5, and section 6 summarizes the paper.

4. Evolution Strategies with Line Search

4.1. Standard ES and L-ES
During a standard evolution strategy, the optimum is sought for by the application of the principles of natural evolution, i.e., recombination, mutation and selection to a set $\mu$ of feasible solutions, the so-called individuals

$$w_k^T = \{x_1, x_2, \ldots, x_N, \sigma_1, \sigma_2, \ldots, \sigma_N\}, \quad k = 1, \ldots, \mu.$$ (1)
In this context, $N$ stands for the problem dimension, and $x_i$ represent the design variables, which define a solution in the search space. The $\sigma_j$ are so called strategy variables, which are used to parameterize the ES during the search. In the above equation, the strategy variables can be identified as the step-sizes for the mutation process. Thus, one individual represents one complete design and contains additional information to steer the optimization algorithm. In the following, we restrict ourselves for the sake of brevity to the case $N = 1$. More information about other parameterizations can be found in [14]. The first step to create a number of $\lambda \geq \mu$ new designs is to exchange or to average the properties of randomly selected parents

$$w'_{ki} = \begin{cases} w_{pi} \text{ or } w_{qi} & \text{(discrete)} \\ 1/2 \cdot (w_{pi} + w_{qi}) & \text{(intermediate)} \end{cases} \text{ for } k = 1, \ldots, \lambda, \quad i = 1, \ldots, N, \quad p, q \sim U(1, \mu).$$ (2)

This procedure is named recombination. The parameters $p, q \sim U(1, \mu)$ represent realizations of random variates underlying the uniform distribution $U(1, \mu)$ and the prime indicates, that the individual has been created by recombination. The discrete and intermediate recombination is named global, if the recombining parents are selected anew for the creation of each vector component $i$. If two parents are chosen and recombination is performed for the whole individual, the procedure is termed local. Then mutation takes place, where small random changes of the individuals are performed. The intensity of the changes is determined by the strategy variables. The process starts with the variation of the mutation step-size

$$\sigma_k^\mu = \sigma_k^\mu \cdot \exp \left( \tau_1 \cdot \mathcal{N}(0, 1) \right), \quad \tau_1 \sim \frac{1}{\sqrt{N}}, \quad k = 1, \ldots, \lambda,$$ (3)

and for the design variables, the mutation scheme reads

$$x_{ki}' = x_{ki}^l + \sigma_k^\mu \cdot \mathcal{N}(0, 1), \quad i = 1, \ldots, N, \quad k = 1, \ldots, \lambda.$$ (4)

In this context $\mathcal{N}(0, 1)$ denotes a standard normal distributed random variate, and the index $i$ indicates, that a new random number is drawn for each component $i$. The described mutation process was developed by [15]. More variants exist; detailed information can be found in [16, 17, 18, 19]. The double prime indicates, that the individual was created by mutation. After having performed the evaluation of the objective function for each individual, the $\mu$ best individuals are selected to become the parents for the next iteration step, which is named generation. In the case of a $(\mu+\lambda)$-selection the $\mu$ best individuals are chosen from the set of the union of parents and offspring, and in the case of a $(\mu, \lambda)$-selection the best $\mu$ individuals are chosen from the set of the offspring only. Recombination, mutation and selection are then repeated until a termination criterion, such as a lower limit for the mutation step-sizes or a maximum number of generations, is fulfilled. The algorithm is summarized as follows:

$$g := 0$$
create the initial population $P_0$

**do while** termination criterion is not fulfilled

- apply recombination to $\mu$ parents and create $\lambda$ offspring
- mutate the offspring
- evaluate the objective function for each individual
- select the new parents for $P_{g+1}$

$$g := g + 1$$

**end do**

For detailed information about evolution strategies it is referred to [14, 16, 17, 18].

Beside the good global search properties and the robustness of evolution strategies, the weak point of the method is the high computational effort. The evolution strategies with line search was thus developed with the following purposes:

- reduction of the objective function calls being necessary to locate the optimum,
- retainance of robustness and good global search properties of the standard ES,
- general applicability and proper operation for all parameterizations of the ES.
For the L-ES an extrapolation factor $\zeta_k$ is defined as a new strategy variable. Thus a parental individual appears now in the form

$$w_k^T = [x_1, x_2, \ldots, x_N, \sigma, \zeta], \quad k = 1, \ldots, \mu. \quad (5)$$

The basic idea of our approach is to save for each individual $k$ the successful mutation vectors $z_k^{(g-1)}, z_k^{(g-2)}, \ldots, z_k^{(g-N_h)}$ from the last $N_h$ past generations. During a normal ES run, in certain time intervals defined by a parameter $N_e$, these vectors are used for the computation of an average success direction (respectively a gradient approximation) per individual

$$d_k = \frac{1}{N_h} \sum_{j=1}^{N_h} \frac{z_k^{(g-j)}}{|z_k^{(g-j)}|}, \quad k = 1, \ldots, \mu. \quad (6)$$

After recombining the design variables, the strategy variables and the gradient approximations, the mutation for all strategy variables is carried out. A line search with a newly developed self-adaptive algorithm is then performed along the success direction of each individual, and extrapolation steps are carried out. The extrapolation step-size $\zeta_k^*$ is computed by multiplying the extrapolation factors $\zeta_k^*$ and the mutation step-size $\sigma_k^*$

$$\zeta_k^* = \zeta_k^0, \quad k = 1, \ldots, \lambda. \quad (7)$$

The offspring are finally created by extrapolation along $x_k^\prime + \zeta_k^* d_k$ and superimposed mutation

$$x_k^\prime = x_k + \zeta_k^* d_k + z_k, \quad k = 1, \ldots, \lambda, \quad (8)$$

where the components of the random vector $z_k$ are calculated according to the scheme of the standard evolution strategy, Eq. (4), via $z_k = \sigma_k^i, N_i(0, 1), i = 1, \ldots, N$. After the mutation process, the standard ES is continued for the next $N_e - 1$ generations. The basic algorithm of the L-ES is summarized below.

$$g := 0$$

create the initial population $P_0$

**do while** termination criterion is not fulfilled

**if** $(g \mod N_e = 0)$ and $(g \geq N_h)$ **then**

compute the average success directions $d_k$ for all $\mu$ parents

apply recombination to $\mu$ parents and create $\lambda$ offspring

mutate the mutation step-sizes and the extrapolation factors

compute an extrapolation step-size $\zeta_k^*$ for all $\lambda$ offspring

extrapolate and mutate the design variables $x_k$ for all $\lambda$ offspring

**else**

apply recombination to $\mu$ parents and create $\lambda$ offspring

mutate the offspring

**end if**

evaluate the objective function for each individual

select the new parents for $P_{g+1}$

update the reservoir of stored successful mutation vectors

$g := g + 1$

**end do**

4.2. Test results

Numerical experiments with a number of standard mathematical benchmark functions are performed to investigate the problem solving capacity of the evolution strategy with line search. The functions are listed in Table 1. As basis for the experiments a (10,70)-ES with global-discrete recombination of the design variables and global intermediate recombination of the mutation step-sizes is considered. The initial design is chosen randomly in the interval $-10 \leq x_i \leq 10, i = 1, \ldots, N$, the initial mutation step-size is set to $\sigma_m = 0.3$, and for the initial extrapolation factor we use a value of $\zeta_m = 2$. The results shown below in figures 1 and 2 present the average of 30 optimization runs, where one optimization task is considered to be solved, if a objective function value of $f_0 = 10^{-8}$ is reached. The problem dimension is $N = 30$. Numerical tests are performed for all benchmark functions to investigate the influence of the parameters $N_h$ and $N_e$ on the performance of the L-ES. These results were published in [20]. The iteration histories, i.e., the relationships between objective function value and the number of objective function evaluations, are shown in figures 1 (a), (b) and 2 (a)-(c) for optimal parameters $N_h$ and $N_e$. The
Table 1: Mathematical benchmark functions.

<table>
<thead>
<tr>
<th>Name</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sphere</td>
<td>$f(x) = \sum_{i=1}^{N} x_i^2$</td>
</tr>
<tr>
<td>ellipsoid</td>
<td>$f(x) = \sum_{i=1}^{N} i \cdot x_i^2$</td>
</tr>
<tr>
<td>Schwefel's ridge</td>
<td>$f(x) = \sum_{i=1}^{N} \left[ \sum_{j=1}^{i} x_j \right]^2$</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>$f(x) = \sum_{i=1}^{N-1} \left[ 100(x_i^2 - x_{i+1}^2 + (x_i - 1)^2 \right]$</td>
</tr>
<tr>
<td>Ackley</td>
<td>$f(x) = 20 + e - \exp \left( \frac{1}{N} \sum_{i=1}^{N} \cos \left( 2\pi x_i \right) \right) - 20 \cdot \exp \left( -\frac{1}{4} \sqrt{\sum_{i=1}^{N} x_i^2} \right)$</td>
</tr>
</tbody>
</table>

The influence of the problem dimension on the L-ES can be obtained from figure 2 (d), where the number of function evaluations to reach $f_0$ depending on the problem dimension is depicted. Note that the problem condition $\kappa$ gets worse with increasing dimension for the ellipsoid and Schwefel's ridge. For the latter function, the condition is $\kappa = 52813$ for $N = 180$ [17]. Compared to the standard ES, the L-ES performs significantly better for the ellipsoid, Schwefel's ridge and the Rosenbrock function. Obviously the new algorithm is able to adapt itself rapidly to the topology of scaled problems. The minimum of the highly multimodal Ackley-function could be located by the L-ES. For isotropic, unscaled problems, (sphere and Ackley-function) the standard ES is able to adapt an optimal mutation step-size which guarantees a maximum progress rate [17]. In this cases the extrapolations during the L-ES slow down the (optimal) adaption and slightly more function evaluations are necessary [20].

Figure 1: Optimization results, part I. Sphere and Ellipsoid: $N_e = 1, N_h = 10$.

*For quadratic problems $f(x) = x^T A x$ with symmetric and positive definite matrix $A \in \mathbb{R}^{N \times N}$ and arbitrary $x \in \mathbb{R}^N$, the problem condition $\kappa$ can be defined as the ratio of the largest and the smallest eigenvalue of $A$. Problems are said to be ill conditioned for $\kappa \gg 1$. 
5. **Optimization of a composite beam**

As a technical application a composite beam shown in figure 3 is considered.
Both the side panels and the connecting panels consist of 20 layers of carbon fiber reinforced epoxy. Each layer has a thickness of 1 mm and the volume fraction of the fibers is 40%. The material properties and the normalized loads can be found in Table 2, where the index $L$ is used for a property in fiber direction and the index $T$ indicates a property orthogonal to the fiber direction.

<table>
<thead>
<tr>
<th>material properties</th>
<th>normalized load</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_L = 825.50$ GPa</td>
<td>$F^1_x = 0.18$</td>
</tr>
<tr>
<td>$E_T = 6269$ GPa</td>
<td>$F^2_x = 0.81$</td>
</tr>
<tr>
<td>$G_{LT} = 2909$ GPa, $\nu_{LT} = 0.3$</td>
<td>$F^3_x = -1.00$</td>
</tr>
</tbody>
</table>

The optimization task is here to maximize the stiffness of the structure by adapting the ply orientation angles $\beta_i \in [-90^\circ, 90^\circ]$, $i = 1, \ldots, 20$, of the side panels and the connecting panels. Since both side panels and both connecting panels are identical, the problem is 40-dimensional. The definition of the ply orientation is shown in Figure 4.

![Figure 4: Coordinate systems for the definition of the ply orientations](image)

The optimization problem is restricted by the failure criteria

$$F^1_c = \max \left\{ \left| \frac{\varepsilon_x}{\varepsilon_x^{+}} \right|, \left| \frac{\varepsilon_y}{\varepsilon_y^{+}} \right|, \left| \frac{\varepsilon_{xy}}{\varepsilon_{xy}^{+}} \right| \right\} < 1,$$

$$F^2_c = \max \left\{ \frac{\sigma_x}{R_x^{+}} , \frac{\sigma_y}{R_y^{+}} , \frac{\sigma_{xy}}{R_{xy}^{+}} \right\} < 1,$$

$$F^3_c = \frac{\sigma_x^2}{R_x^{+}R_x^{-}} + \frac{\sigma_y^2}{R_y^{+}R_y^{-}} - \frac{\sigma_x\sigma_y}{\sqrt{R_x^{+}R_y^{-}}R_y^{+}R_y^{-}} + \frac{\sigma_{xy}^2}{R_{xy}^{+}R_{xy}^{-}} + \frac{1}{R_x^{+}} + \frac{1}{R_y^{+}} \cdot \sigma_x + \frac{1}{R_x^{-}} + \frac{1}{R_y^{-}} \cdot \sigma_y < 1.$$  

Eq. (9) is the maximum strain criterion, Eq. (10) is the maximum stress criterion and Eq. (11) is the Tsai-Wu criterion. The variable $\varepsilon$ in Eq. (9) refers to a critical strain value and the variable $R$ in Eqs. (10), (11) is a critical strength. The index $(+)$ characterizes a pressure load, and $(-)$ indicates a tensile load. If, like in Eqs. (9) and (10), both indices are used together, the critical value must be chosen according to the type of load. By using the elastic strain energy

$$2E_p(\beta) = \int_{(V)} \sigma(\beta)^T \varepsilon(\beta) \, dV$$

as indirect measure for the structural stiffness and applying the above given restrictions in the form

$$F_c = \max \left\{ F^1_c, F^2_c, F^3_c \right\},$$
the objective function to be minimized reads
\[
\min_{\beta} f(\beta) = \int (V) \sigma^T \varepsilon \, dV + \gamma \cdot \left\{ \begin{array}{ll}
(\bar{F}_c - 1)^2 & \text{if } \bar{F}_c \geq 1 \\
0 & \text{else}
\end{array} \right.
\]  
(14)

During a first optimization, the stiffness of the composite beam is maximized without restrictions. For the optimization with restrictions, a penalty parameter \(\gamma = 10^{20}\) is used. The optimization was carried out with a (10,70)-ES and a (10,70)-L-ES with global discrete recombination of the design variables and local intermediate recombination of the strategy variables. For the L-ES, the parameters \(N_b = 10\) and \(N_c = 2\) were used. The results of the optimization runs can be obtained from figure 5, which shows the fiber orientations for each layer of the side panels and the connecting panels.

(a) side panels  
(b) connecting panels

Figure 5: Optimal ply orientations

For the side panels, the ply orientations in the layers 1, 2, 3, 4 and 20, 19, 18, 17 are similar. The 90° orientations at the outer layers of the side panels reduce the deflections due to the bending moment about the x-axis and provide an optimal stiffness for this part of the loading regime. A detailed analysis showed, that the orientations of the inner layers are caused by the outer shape of the side panels. The resulting ply orientations in the connecting panels of about 0° cause a high stiffness of the beam if bended about the z-axis. Investigations of the objective function revealed, that the optimization problem is highly multimodal. Thus both methods required some restarts to reach the above discussed optimal solution. There is no proof, that the global optimum is obtained, but since no other solution could be found with numerous other parameterizations of both optimization methods and the results allow a meaningful mechanical interpretation, we assume, that the global optimum was located. Figure 6 shows exemplarily the iteration histories of ES and L-ES for the unrestricted optimization problem. Similar to the results obtained during the optimization of the multimodal ACKLEY function, both methods need approximately the same number of generations.

Figure 6: Iteration histories for the ES and the L-ES.
6. Conclusions
A new variant of evolution strategies termed the evolution strategies with line search is developed. The L-ES can be used with the full range of possible parameterizations of the standard evolution strategies, and it is not restricted to special problem types. Numerical tests with a set of standard mathematical benchmark functions are performed. The results show, that, compared to the ES, the L-ES needs significantly less function evaluations for scaled optimization problems. At the same time the robustness and the good global search properties of the ES are maintained. An application from structural mechanics underlines, that the L-ES can be applied successfully to practical optimization problems.

7. Acknowledgements
The support of the DFG-Graduiertenkolleg 828 Micro-Macro-Interactions in Structured Media and Particle Systems is gratefully acknowledged. The work was also supported by the European Commission in the frame of the research project COMO – COmpetence in MObility.

References


