Chapter 5
Iterated Local Search

5.1 Introduction

Hybridization has developed to an effective strategy in algorithm design. Hybrid algorithms can become more efficient and more effective than their *native* counterparts. This observation holds true for many problem classes, in particular in optimization, where hybrid techniques of meta-heuristics and local search are often called hybrid meta-heuristics. In this chapter, we show how Powell's conjugate gradient search, which is a fast and powerful black box optimization strategy for convex problems, can be integrated into an ES [1]. Further, we show how to employ a specialized step size adaptation technique that allows to guide the optimization process and to escape from local optima that Powell's method may successively find.

5.2 Iterated Local Search

Iterated local search (ILS) is based on a simple but successful idea. Instead of repeating local search and starting from initial solutions like restart approaches do, ILS begins with a solution $\mathbf{x}$ and successively applies local search and perturbation of the local optimal solution $\hat{\mathbf{x}}$. This procedure is repeated iteratively until a termination condition is fulfilled. Algorithm 1 shows the pseudocode of the ILS approach. Initial solutions should use as much information as possible to be a good starting point for local search. Most local search operators are deterministic. Consequently, the perturbation mechanism should introduce non-deterministic components to explore the solution space. The perturbation mechanism performs global random search in the space of local optima that are approximated by the local search method. Blum and Roli [2] point out that the balance of the perturbation mechanism is quite important. Perturbation must be strong enough to allow the escape from basins of attraction, but weak enough to exploit knowledge from previous iterations. Otherwise, the ILS will

become a simple restart strategy. The acceptance criterion of Line 6 may vary from *always accept* to *only accept in case of improvement*. Approaches like simulated annealing may be adopted.

Algorithm 1 Iterated Local Search

1: initialize solution **x**
2: produce **x̂** with local search
3: **repeat**
4:     perturbation of **x**
5:     produce **x̂** with local search
6:     apply acceptance criterion
7: **until** termination condition

There are many examples in literature for the successful application of ILS variants on combinatorial optimization problems. A survey of ILS techniques has been presented by Lourenco et al. [3]. The authors also provide a comprehensive introduction [4] to ILS. A famous combinatorial instance, many ILS methods have been developed for, is the traveling salesperson problem. Stützle and Hoos [5] introduced an approach that combines restarts with a specific acceptance criterion to maintain diversity for the TSP, while Katayama and Narihisa [6] use a perturbation mechanism that combines the heuristic 4-opt with a greedy method. Stützle [7] uses an ILS hybrid to solve the quadratic assignment problem. The technique is enhanced by acceptance criteria that allow moves to worse local optima. Furthermore, population-based extensions are introduced. Duarte et al. [8] introduce an ILS heuristic for the problem of assigning referees to scheduled games in sports based on greedy search. Our perturbation mechanism is related to their approach. Preliminary work on the adaptation of the perturbation algorithm has been applied by Mladenovic et al. [9] for variable neighborhood search and tabu search by Glover et al. [10].

5.3 Powell's Conjugate Gradient Method

The hybrid ILS variant introduced in this chapter is based on Powell's optimization method. Preliminary experiments revealed the efficiency of Powell's method in comparison to continuous evolutionary search methods. However, in the experimental section, we will observe that Powell's method can get stuck in local optima in multimodal solution spaces. An idea similar to the hybridization of local search has been presented by *Griewank* [11], who combines a gradient descent method with a deterministic perturbation term.

Powell's method belongs to the class of direct search methods, i.e., no first or second order derivatives are required. It is based on conjugate directions and is similar to line search. The idea of line search is to start from search point $\mathbf{x} \in \mathbb{R}^N$ along a direction $\mathbf{d} \in \mathbb{R}^N$, so that $f(\mathbf{x} + \lambda_t \mathbf{d})$ is minimized for a $\lambda_t \in \mathbb{R}^+$. Powell's

method [12, 13] adapts the directions according to a gradient-like information from the search.

Algorithm 2 Powell's Method

1: **repeat**
2: **for** $t = 1$ to N **do**
3:      find λ_t that minimizes $f(\mathbf{x}_{t-1} + \lambda_t \mathbf{d}_t)$
4:      set $\mathbf{x}_t = \mathbf{x}_{t-1} + \lambda_t \mathbf{d}_t$
5:      **for** $j = 1$ to $N - 1$ **do**
6:          update vectors $\mathbf{d}_j = \mathbf{d}_{j+1}$
7:      **end for**
8:      set $\mathbf{d}_N = \mathbf{x}_N - \mathbf{x}_0$
9:      find λ_N that minimizes $f(\mathbf{x}_N + \lambda_N \mathbf{d}_N)$
10:     set $\mathbf{x}_0 = \mathbf{x}_0 + \lambda_N \mathbf{d}_N$
11: **end for**
12: **until** termination condition

It is based on the assumption of a quadratic convex objective function $f(\mathbf{x})$

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{H}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c. \tag{5.1}$$

with Hessian matrix $\mathbf{H}$. Two directions $\mathbf{d}_i, \mathbf{d}_j \in \mathbb{R}^N$, $i \neq j$ are mutually conjugate, if

$$\mathbf{d}_i^T \mathbf{H}\mathbf{d}_j = 0 \tag{5.2}$$

holds with mutual conjugate directions that constitute a basis of the solution space $\mathbb{R}^N$. Let $\mathbf{x}_0$ be the initial guess of a minimum of function f. In iteration t, we require an estimation of the gradient $\mathbf{g}_t = \mathbf{g}(\mathbf{x}_t)$. Let $t = 1$ and let $\mathbf{d}_t = -\mathbf{g}_t$ be the steepest descent direction. For $t > 1$, Powell applies the equation

$$\mathbf{d}_t = -\mathbf{g}_t + \beta_t \mathbf{d}_{t-1}, \tag{5.3}$$

with the Euclidean vector norms

$$\beta_t = \frac{\|\mathbf{g}_t\|^2}{\|\mathbf{g}_{t-1}\|^2}. \tag{5.4}$$

The main idea of the conjugate direction method is to search for the minimal value of $f(\mathbf{x})$ along direction $\mathbf{d}_t$ to obtain the next solution $\mathbf{x}_{t+1}$, i.e., to find the λ that minimizes

$$f(\mathbf{x}_t + \lambda_t \mathbf{d}_t). \tag{5.5}$$

For a minimizing λ_t, set

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \lambda_t \mathbf{d}_t. \tag{5.6}$$

Algorithm 2 shows the pseudocode of the conjugate gradient method that is the basis of Powell's strategy. In our implementation, the search for λ_t is implemented with line search. For a more detailed introduction, we refer to the depiction of Powell [12] and Schwefel [14].

At first, we analyze Powell's method on the optimization test suite (cf. Appendix A). Solutions are randomly initialized in the interval $[-100, 100]^N$. Each experiment is repeated 30 times. Powell's method terminates, if the improvement from one to the next iteration is smaller than $\phi = 10^{-10}$ with comma selection, or if the optimum is found with accuracy $f_{stop} = 10^{-10}$. As Powell's method is a convex optimization technique, we expect that only the unimodal problems can be solved. Table 5.1 confirms these expectations. On unimodal functions, Powell's method is exceedingly fast. On the *Sphere* problem with $N = 10$, a budget of only 101.7 fitness function evaluations in mean is sufficient to approximate the optimum. These fast approximation capabilities can also be observed on problems *Doublesum* and *Rosenbrock*, also for higher dimensions, i.e. $N = 30$.

The results also show that Powell's method is not able to approximate the optima of the multimodal function *Rastrigin*. On the *easier* multimodal function *Griewank*, the random initializations allow to find the optimum in some of the 30 runs. The fast convergence behavior on convex function parts motivates to perform local search as operator in a global evolutionary optimization framework. It is the basis of the Powell ES that we will analyze in the following.

Table 5.1 Experimental comparison of Powell's method on the test problems with $N = 10$ and $N = 30$ dimensions

	Best	Median	Worst	Mean	Dev	#
$N = 10$						
f_{Sp}	100	102	102	101.7	0.67	30
f_{Dou}	91	92	92	91.8	0.42	30
f_{Ros}	2,947	4,617	12,470	5,941.87	3,353.14	24
f_{Ras}	–	–	–	–	–	0
f_{Gri}	329	329	329	329	0	1
f_{Kur}	–	–	–	–	–	0
$N = 30$						
f_{Sp}	299	302	302	301.3	1.05	30
f_{Dou}	291	291.5	292	291.5	0.52	30
f_{Ros}	14,888	33,315	59,193	36,455.85	16,789.41	21
f_{Ras}	–	–	–	–	–	0
f_{Gri}	904	997	1,001	967.33	54.88	3
f_{Kur}	–	–	–	–	–	0

Best, *median*, *worst*, *mean*, and *dev* provide statistical information about the number of fitness function evaluations of 30 runs until the difference between the fitness of the best solution and the optimum is smaller than $f_{stop} = 10^{-10}$. Parameter # states the number of runs that reached the optimum.

5.4 Powell Evolution Strategy

The Powell ES [1] presented in this section is based on four key concepts, each focusing on typical problems in real-valued solution spaces. Powell's method is a fast direct optimization method, in particular appropriate for unimodal fitness landscapes. It is integrated into the optimization process using ILS, in order to prevent Powell's method from getting stuck in local optima. ILS approach is based on the successive repetition of Powell's conjugate gradient method as local search technique and a perturbation mechanism. A population of candidate solutions is employed for exploration similar to evolution strategies. The strength of the ILS perturbation is controlled by means of an adaptive control mechanism. In case of stagnation, the mutation strength is *increased*, in order to leave local optima, and *decreased* otherwise.

Algorithm 3 shows the pseudocode of the Powell ES. At the beginning, μ solutions $\mathbf{x}_1, \ldots, \mathbf{x}_\mu \in \mathbb{R}^N$ are randomly initialized and optimized with the strategy of Powell. In an iterative loop λ, offspring solutions $\mathbf{x}_1, \ldots, \mathbf{x}_\lambda$ are produced by means of Gaussian mutations with the global mutation strength σ by

$$\mathbf{x}'_j = \mathbf{x}_j + \mathbf{z}_j, \tag{5.7}$$

with

$$\mathbf{z}_j \sim (\sigma_1 \mathcal{N}(0, 1), \ldots, \sigma_N \mathcal{N}(0, 1))^T. \tag{5.8}$$

Afterwards, each solution $\mathbf{x}'_j$ is locally optimized with the strategy of Powell, leading to $\hat{\mathbf{x}}'_j$ for $j = 1, \ldots, \lambda$. After λ solutions have been produced this way, the μ-best are selected according to their fitness with comma selection. Then, we apply global recombination, i.e., the arithmetic mean $\langle \hat{\mathbf{x}}_t \rangle$ at generation t of all selected solutions $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_\mu$ is computed. The fitness of this arithmetic mean is evaluated and compared to the fitness of the arithmetic mean of the last generation $t-1$. If the search stagnates, i.e., if the condition

$$|f(\langle \hat{\mathbf{x}} \rangle_t) - f(\langle \hat{\mathbf{x}} \rangle_{t-1})| < \theta \tag{5.9}$$

becomes true, the mutation strength is increased via

$$\sigma = \sigma \cdot \tau \tag{5.10}$$

with $\tau > 1$. Otherwise, the mutation strength σ is decreased by multiplication with $1/\tau$.

An increasing mutation strength σ allows to leave local optima. Powell's method drives the search into local optima, and the outer ILS performs a search within the *space of local optima* controlling the perturbation strength σ. A decrease of step size σ lets the algorithm converge to the local optimum in a range defined by σ. This technique seems to be in contraposition to the 1/5th success rule by Rechenberg [15]. Running a simple $(1 + 1)$-ES with isotropic Gaussian mutations

---

Algorithm 3 Powell ES

1: initialize μ solutions
2: apply Powell's method
3: **repeat**
4:    **for** $j = 1$ **to** λ **do**
5:       mutate solution $\mathbf{x}_j$
6:       apply Powell's method
7:    **end for**
8:    select μ-best solutions
9:    **if** fitness improvement $< \theta$ **then**
10:       $\sigma = \sigma \cdot \tau$
11:    **else**
12:       $\sigma = \sigma / \tau$
13:    **end if**
14: **until** termination condition

---

and constant mutation steps σ, the optimization process will become very slow after a few generations. Rechenberg's rule adapts the mutation strengths in the opposite kind of way. If the ratio g/G of successful generations g after G generations is larger than 1/5th, the step size should be increased. The increase is reasonable, because bigger steps towards the optimum are possible, while small steps would be a waste of time. If the success ratio is less than 1/5th, the step size should be decreased. This rule is applied every G generations. The goal of Rechenberg's approach is to stay in the *evolution window* guaranteeing nearly optimal progress. Optimal progress is problem-dependent and can be stated theoretically on artificial functions [16].

However, in our approach the strategy of Powell approximates local optima, not the evolution strategy. The step control of the Powell ES has another task: leaving local optima, when the search stagnates. Basins of attractions can be left because of the increasing step size. Hence, the probability of finding the global optimum is larger than 0. With this mechanism, also the global optimum may be left again. But if the vicinity of the optimum has been reached, it is probable that it will be successively reached again. The problem that the global optimum may be left, if not recognized, can be compensated by saving the best found solution in the course of the optimization process.

5.5 Experimental Analysis

In the following, we will experimentally analyze the Powell ES on a set of test problems, cf. Appendix A. Again, initial solutions are generated in the interval $[-100, 100]^N$, and the step sizes are set to $\sigma_{\text{init}} = 1.0$. Each experiment is repeated 30 times. For the Powell ES, we employ the settings $\lambda = 8$ and $\mu = 2$. Each solution is mutated and locally optimized with Powell's method. Again, Powell's method terminates, if the improvement from one to the next iteration is smaller than $\phi = 10^{-10}$, or if the optimum is found with accuracy $f_{\text{stop}} = 10^{-10}$. If the

search on the ILS level stagnates, i.e., if the achieved improvement is smaller than θ, the mutation strength is increased with mutation parameter $\tau = 2.0$. We allow a maximal budget of $\text{ffe}_{max} = 2.0 \times 10^6$ fitness function evaluations.

Table 5.2 shows the results of the analysis of the Powell ES on the test problems with $N = 10$ and $N = 30$ dimensions. The results have shown that Powell's method is very fast on unimodal problems. Of course, the Powell ES shows the same capabilities and approximates the optimum in the first Powell-run on the *Sphere* problem and *Doublesum*. We have already observed that Powell gets stuck in local optima of multimodal problems (e.g. *Rastrigin*). The Powell ES perturbates a solution, when getting stuck, and applies Powell's method again with the perturbation mechanism of Eq. (5.10). The results show that the iterated application of Powell's method in each generation allows to approximate the global optimum, also on *Rastrigin*. The Powell ES is able to approximate the optimum in comparison to its counterpart without ILS.

It convergences significantly faster than the CMSA-ES (cf. Chap. 2). A statistically significant superiority of the Powell ES can also be observed on *Griewank*. On *Rosenbrock*, no superiority of any of the two algorithms can be reported. Although the worst runs of the Powell ES cause a fitness deterioration in mean, the best runs are still much faster than the best runs of the CMSA-ES. The CMSA-ES is more robust with smaller standard deviations, but does not offer the potential to find the optimal solution that fast. A similar behavior can be observed on the test problems with $N = 30$ dimensions, see the lower part of Table 5.2. The CMSA-ES takes about 17 times more evaluations. This also holds true for the other unimodal test problems, where the Powell ES is superior. On the multimodal test problems in higher dimensions, similar results as for $N = 10$ can be observed. The Powell ES

Table 5.2 Experimental analysis of the Powell ES on the test problems with $N = 10$ and $N = 30$ dimensions

	Best	Median	Worst	Mean	Dev	#
$N = 10$						
f_{Sp}	99	100	153	105.1	1.6e1	30
f_{Dou}	89	92	178	108.6	3.6e1	30
f_{Ros}	3,308	5,074.5	29,250	7,772.6	7.8e3	30
f_{Ras}	2,359	14,969.5	38,550	15,682.5	9.1e3	30
f_{Gri}	477	1,506	6,572	2,240	2.2e3	30
f_{Kur}	4,300	196,218	316,528	165,325	9.9e4	30
$N = 30$						
f_{Sp}	290	295.5	299	295.3	2.83	30
f_{Dou}	283	286.5	479	305.6	6.0e1	30
f_{Ros}	25,363	61,768	385,964	95,339	1.0e5	30
f_{Ras}	58,943	78,537.5	191,489	102,429	4.4e4	30
f_{Gri}	971	5,994.5	20,913	9,629.6	7.6e3	30
f_{Kur}	>2.0e6	>2.0e6	>2.0e6	>2.0e6	–	0

The figures show the number of fitness function evaluations until the difference between the fitness of the best solution and the fitness of the optimum is smaller than $f_{stop} = 10^{-10}$. This termination condition has been reached in every run except on f_{Kur} with $N = 30$.
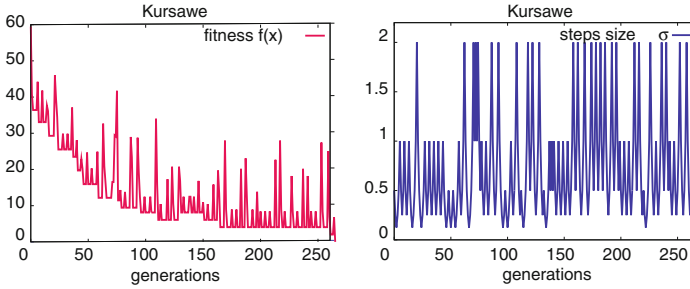
Fig. 5.1 Development of fitness and step sizes on the multimodal problem *Kursawe* employing $N = 10$. When the search gets stuck in local optima, the perturbation mechanism increases σ and lets the Powell ES escape from basins of attraction [1]

is statistically better on *Rastrigin*. The CMSA-ES's mean and median are better on *Rosenbrock* and *Griewank*. On *Kursawe*, the optimum has been found in every run for $N = 10$, but in no run for $N = 30$.

Figure 5.1 shows fitness curves and step sizes of typical runs on the multimodal problem *Kursawe* with $N = 10$. It can be observed that the search successively gets stuck. But the perturbation mechanism always allows to leave the local optima again. When the search gets stuck in a local optimum, the strategy increases σ until the local optimum is successfully left, and a better local optimum is found. The approach moves from one local optimum to another controlling σ, until the global optimum is found. The fitness development reveals that the search has to accept worse solutions to approximate the optimal solution. The figures confirm the basic idea of the algorithm. ILS controls the global search, while Powell's method drives the search into local optima. Frequently, the hybrid is only able to leave local optima by controlling the strength σ of the Gaussian perturbation mechanism. ILS conducts a search in the space of local optima.

5.6 Perturbation Mechanism and Population Sizes

For deeper insights into the perturbation mechanism and the interaction with population sizes, we conduct further experiments on the multimodal problem *Rastrigin* with $N = 30$, where the Powell ES has shown successful results. The strength of the perturbation mechanism plays an essential role for the ILS. In case of stagnation, the step size is increased as described in Eq. (5.10) with $\tau > 1$ to let the search escape from local optima. Frequently, a successive increase of the perturbation strength is necessary to prevent stagnation. In case of an improvement, the step size is decreased with $\tau < 1$. The idea of the step size reduction is to prevent the search process from jumping over promising regions of the solution space. In the following, we ana-

Table 5.3 Analysis of the Powell ES perturbation parameter τ and the population sizes on *Rastrigin* with $N = 30$ using the same initial settings, performance measure, and termination condition like in the previous experiments

(μ, λ)	Best	Median	Worst	Best	Median	Worst
	$\tau = 1.5$			$\tau = 2.0$		
(1, 4)	53,913	92,544	130,675	31,686	78,214	121,170
(2, 8)	56,074	100,835	143,642	65,540	112,643	242,983
(4, 16)	149,350	162,083	210,384	77,481	117,972	163,693
(8, 32)	156,517	295,457	370,320	193,259	209,725	244,325
	$\tau = 5.0$			$\tau = 10.0$		
(1, 4)	53,465	105,513	406,495	$>2 \times 10^9$	$>2 \times 10^9$	$>2 \times 10^9$
(2, 8)	48,274	104,461	285,651	32,773	680,363	1,473,097
(4, 16)	67,241	103,142	202,447	52,991	208,088	338,922
(8, 32)	109,820	189,676	221,069	123,838	309,169	802,285

lyze the perturbation mechanism and the population sizes on *Rastrigin*. We try to determine useful parameter settings for τ and for population parameters μ and λ.

Table 5.3 shows the corresponding results. The best result has been achieved with $\tau = 2.0$ and population sizes $(1, 4)$. Also the best median has been achieved with this setting, while the second best has been achieved with $\tau = 1.5$ and population sizes $(1, 4)$. With parameter setting $\tau = 10.0$, the Powell ES achieves a satisfying best solution, but the variance of the results is high. The worst solution is comparably bad. In general, the results for $\tau = 10.0$ are quite weak, for $(1, 4)$ the algorithm does not converge within reasonable time. For low mutation strengths, the best results can be observed for small population sizes. In turn, for higher mutation strengths, i.e., $\tau = 5.0$, larger population sizes are necessary to compensate the explorative effect. Further experiments on other problems led to the decision that a $(2, 8)$-Powell ES is a good compromise between exploration and efficiency, while a $(4, 16)$-Powell ES is a rather conservative, but stable choice with reliable results.

5.7 Conclusions

Combining the world of local search with the world of global evolutionary optimization is a promising undertaking. It reflects the original idea of evolutionary computation. If we do not know anything about the problem, evolutionary algorithms are an appropriate choice. In multimodal fitness landscapes, we typically know nothing about the *landscape of local optima*. The Powell ES only assumes that attractive local optima lie closely together. Hence, the search might jump from one basin of attraction to a neighbored one. To move into local optima, Powell's method turns out to be fairly successful. Furthermore, the adaptation of the perturbation strength is a natural enhancement in real-valued solution spaces. A population-based

implementation allows to run multiple Powell searches in parallel and to achieve a crucial speedup in distributed computing environments.

References

1. O. Kramer, Iterated local search with Powell's method: a memetic algorithm for continuous global optimization. Memetic Comput. **2**(1), 69–83 (2010)
2. C. Blum, M. Aguilera, A. Roli, M. Sampels, Hybrid Metaheuristics: an Introduction. In Hybrid Metaheuristics (Springer, Berlin, 2008), pp. 1–30
3. H.R. Lourenço, O. Martin, T. Stützle, *Iterated Local Search*, eds. by F. Glover, G. Kochenberger. In Handbook of Metaheuristics, (Kluwer Academic, Norwell, 2003), pp. 321–352
4. H.R. Lourenço, O. Martin, T. Stützle, A beginner's introduction to iterated local search, in *Proceedings of the Fourth Metaheuristics Conference*, vol. 2 (Porto, Portugal, 2001), pp. 1–6
5. T. Stützle, H.H. Hoos, Analyzing the run-time behaviour of iterated local search for the TSP, in *Proceedings III Metaheuristics International Conference* (Kluwer Academic Publishers, Norwell, 1999)
6. K. Katayama, H. Narihisa, Iterated local search approach using genetic transformation to the traveling salesman problem, in *Proceedings of the 1st Conference on Genetic and Evolutionary Computation (GECCO) ACM*, New York, USA, 1999
7. T. Stützle, Iterated local search for the quadratic assignment problem. Eur. J. Oper. Res. **174**(3), 1519–1539 (2006)
8. A.R. Duarte, C.C. Ribeir, A hybrid ILS heuristic to the referee assignment problem with an embedded MIP strategy, *Hybrid Metaheuristics* (2007), pp. 82–95, http://link.springer.com/chapter/10.1007%2F978-3-540-75514-2_7
9. N. Mladenovic, P. Hansen, Variable neighborhood search. Comput. Oper. Res. **24**, 1097–1100 (1997)
10. F. Glover, M. Laguna, *Tabu Search* (Springer, Berlin 1997)
11. A. Griewank, Generalized decent for global optimization. JOTA **34**, 11–39 (1981)
12. M. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives. Comput. J. **7**(2), 155–162 (1964)
13. M.J.D. Powell, Restart procedures for the conjugate gradient method. Math. Program. **V12**(1), 241–254 (1977)
14. H.-P. Schwefel, *Evolution and Optimum Seeking. Sixth-Generation Computer Technology* (Wiley, New York, 1995)
15. I. Rechenberg, *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (Frommann-Holzboog, Stuttgart, 1973)
16. H.-G. Beyer, *The Theory of Evolution Strategies* (Springer, Berlin, 2001)