Fast Blackbox Optimization: Iterated Local Search and the Strategy of Powell

Oliver Kramer

Algorithm Engineering Report
TR09-2-003
Feb. 2009
ISSN 1864-4503
Fast Blackbox Optimization: Iterated Local Search and the Strategy of Powell

Oliver Kramer
Computational Intelligence Group
Department of Computer Science
Dortmund University of Technology, Germany
Email: oliver.kramer@tu-dortmund.de

Abstract—In combinatorial solution spaces Iterated Local Search turned out to be exceptionally successful. The question arises: is the Iterated Local Search heuristic also able to improve the optimization process in real-valued solution spaces? This paper introduces a hybrid meta-heuristic based on Iterated Local Search and Powell’s optimization method combined with elements from stochastic search in real-valued solution spaces. The approach is analyzed experimentally. It turns out that the Iterated Local Search Hybrid is significantly faster than state-of-the-art evolutionary approaches and behaves more robust than the strategy of Powell in multimodal fitness landscapes.

I. INTRODUCTION

Recent results have shown that the hybridization between meta-heuristics and local search techniques turn out to be exceptionally successful — in particular in combinatorial and discrete solution spaces [10], [18]. Interestingly, for real-valued solution spaces not many results have been reported so far. In this paper we introduce a hybrid meta-heuristic that is based on Powell’s method and Iterated Local Search (ILS). First, we will introduce the concept of hybridization in general and the ILS concept in Section II. Section III introduces the ILS-Powell hybrid starting with the strategy of Powell. Section IV provides an experimental evaluation of the proposed approach and concentrates on parameter settings.

II. HYBRID METAHEURISTICS AND ITERATED LOCAL SEARCH

Iterated Local Search belongs to the class of hybrid meta-heuristics. Before we introduce the ILS approach, we give a brief overview of hybrid approaches.

A. Hybrid Meta-Heuristics

Search algorithms can be divided into two categories: exact techniques and heuristics. Exact algorithms find local optimal solutions with great success, but the runtime deteriorates rapidly with the size of the problem dimension. Heuristics and meta-heuristics usually approximate the solution on the basis of stochastic components and do not find the optimum in every case. But their runtime on large problem instances is much more acceptable. A meta-heuristic is a generic design pattern for heuristic algorithm which has to be specified. One of the most important advantages of meta-heuristics is their applicability to a huge number of problems. Even if no knowledge about the problem and the solution space characteristics is available, meta-heuristics can be applied with their stochastic operators. Meta-heuristics like evolutionary algorithms, particle swarm optimization or artificial immune systems have proven well as successful and robust optimization methods within the last decades.

The hybridization of meta-heuristics with local search methods is motivated by the combination of the advantages of the exact and the heuristic techniques. The success of hybridization is reflected by an increasing number of publications in this research area and the foundation of international conferences and workshops like the HM – Hybrid meta-heuristics workshop or the Workshop on Mathematical Contributions to meta-heuristics. In the case of combinatorial solution spaces exact methods like integer linear programming, dynamic programming approaches [1] or branch-and-bound methods [8] are frequently combined with evolutionary algorithms. In numerical solution spaces direct search methods like pattern search [6], simplex search [11], Rosenbrock’s [16] or Powell’s method [13] can be used.

An important design decision for hybrid techniques concerns the way of information interchange between its components. In which order shall the components work together, which information is shared, and when? Can general hybridization rules be derived from theory or experiments? Talbi [19] and Raidl [14] proposed a taxonomy of hybrid metaheuristics, see figure 1.
In their taxonomy a relay or sequential hybrid is a simple successive execution of two or more algorithmic components. The main idea is: A stochastic method might pre-optimize coarsely while the local search performs fine-tuning and approximation of local optima. The coevolutionary or concurrent hybrid is a nested approach. Typically, a local search method is embedded into an evolutionary optimizer: In each iteration the local search optimizes the offspring solutions until a predefined termination condition is fulfilled. Information is passed alternately between the components in the concurrent approach. The local search method might have an own termination condition that can be specified by the embedding optimizer.

B. Iterated Local Search

Iterated Local Search is based on a simple, but successful idea. Instead of simply repeating local search starting from an initial solution like restart-approaches do, ILS optimizes solution \( s \) with local search, perturbs the local optimal solution \( \hat{s} \) and applies local search again. This procedure is repeated iteratively until a termination condition is met. Figure 2 shows the pseudo-code of the ILS approach. Initial solutions should employ as much information as possible to be a fairly 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 some kind of global random search in the space of local optima – that are approximated by the local search method. Blum et al. [4] point out that the balance of the perturbation mechanism is quite important. The perturbation must be strong enough to allow the escape from basins of attraction, but low enough to exploit knowledge from previous iterations. Otherwise, the ILS will become a simple restart strategy. The acceptance criterion of line 7 may vary from always accept to only in case of improvement. Approaches like simulated annealing may be adopted.

III. The ILS-Powell-Hybrid

Our hybrid ILS variant uses Powell’s optimization method.

<table>
<thead>
<tr>
<th>Start</th>
<th>( s \leftarrow \text{generate initial solution;} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{s} \leftarrow \text{localsearch (s);} )</td>
</tr>
<tr>
<td>Repeat</td>
<td>( \hat{s}' \leftarrow \text{perturbation (s);} )</td>
</tr>
<tr>
<td></td>
<td>( \hat{s}' \leftarrow \text{localsearch (s'}); )</td>
</tr>
<tr>
<td></td>
<td>( \hat{s} \leftarrow \text{ApplyAcceptanceCriterion (s', \hat{s});} )</td>
</tr>
<tr>
<td>Until termination condition</td>
<td>End</td>
</tr>
</tbody>
</table>

Fig. 2. Pseudocode of the ILS method

In their taxonomy a relay or sequential hybrid is a simple successive execution of two or more algorithmic components. The main idea is: A stochastic method might pre-optimize coarsely while the local search performs fine-tuning and approximation of local optima. The coevolutionary or concurrent hybrid is a nested approach. Typically, a local search method is embedded into an evolutionary optimizer: In each iteration the local search optimizes the offspring solutions until a predefined termination condition is fulfilled. Information is passed alternately between the components in the concurrent approach. The local search method might have an own termination condition that can be specified by the embedding optimizer.

II. Iterated Local Search

Iterated Local Search is based on a simple, but successful idea. Instead of simply repeating local search starting from an initial solution like restart-approaches do, ILS optimizes solution \( s \) with local search, perturbs the local optimal solution \( \hat{s} \) and applies local search again. This procedure is repeated iteratively until a termination condition is met. Figure 2 shows the pseudo-code of the ILS approach. Initial solutions should employ as much information as possible to be a fairly 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 some kind of global random search in the space of local optima – that are approximated by the local search method. Blum et al. [4] point out that the balance of the perturbation mechanism is quite important. The perturbation must be strong enough to allow the escape from basins of attraction, but low enough to exploit knowledge from previous iterations. Otherwise, the ILS will become a simple restart strategy. The acceptance criterion of line 7 may vary from always accept to only in case of improvement. Approaches like simulated annealing may be adopted.

Fig. 2. Pseudocode of the ILS method

In their taxonomy a relay or sequential hybrid is a simple successive execution of two or more algorithmic components. The main idea is: A stochastic method might pre-optimize coarsely while the local search performs fine-tuning and approximation of local optima. The coevolutionary or concurrent hybrid is a nested approach. Typically, a local search method is embedded into an evolutionary optimizer: In each iteration the local search optimizes the offspring solutions until a predefined termination condition is fulfilled. Information is passed alternately between the components in the concurrent approach. The local search method might have an own termination condition that can be specified by the embedding optimizer.

A. The Strategy of Powell

The classical non-evolutionary optimization methods for continuous problems can mainly be classified into direct, gradient and Hessian search methods. The direct methods determine the search direction without using a derivative [17]. Lewis, Torzcon und Trosset [9] give an overview of direct search methods. Pattern search methods [6] examine the objective function with a pattern of points which lie on a rational lattice. Simplex search [11] is based on the idea that a gradient can be estimated with a set of N+1 points, i.e. a simplex. Direct search methods like Rosenbrock’s [16] and Powell’s [13] collect information about the curvature of the objective function during the course of the search. If the derivatives of a function are available, the gradient and Hessian methods can be applied. Gradient methods take the first derivative of the function into account, while the Hessian methods also compute the second derivative. A successful example is the Quasi-Newton method [5]. It searches for the stationary point of a function, where the gradient is 0. Quasi-Newton estimates the Hessian matrix analyzing successive gradient vectors.

<table>
<thead>
<tr>
<th>Start</th>
<th>( i = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeat</td>
<td>( \text{Set } p_0 = x_i; )</td>
</tr>
<tr>
<td></td>
<td>( \text{For } k = 1 \text{ To } N )</td>
</tr>
<tr>
<td></td>
<td>( \gamma_k \text{ that minimize } f(p_{k-1} + \gamma_k u_k); )</td>
</tr>
<tr>
<td></td>
<td>( \text{Set } p_k = p_{k-1} + \gamma_k u_k; )</td>
</tr>
<tr>
<td>Next</td>
<td>( \text{Next } i = i + 1; )</td>
</tr>
<tr>
<td></td>
<td>( \text{For } j = 1 \text{ To } N - 1; )</td>
</tr>
<tr>
<td></td>
<td>( \text{Update vectors } u_i \text{ by setting } u_j = u_{j+1}; )</td>
</tr>
<tr>
<td>Next</td>
<td>( \text{Next } )</td>
</tr>
<tr>
<td></td>
<td>( \text{Set } u_N = p_n - p_0; )</td>
</tr>
<tr>
<td></td>
<td>( \gamma \text{ that minimizes } f(p_0 + \gamma u_N); )</td>
</tr>
<tr>
<td></td>
<td>( \text{Set } x_i = p_0 + \gamma u_N; )</td>
</tr>
<tr>
<td>Until termination condition</td>
<td>End</td>
</tr>
</tbody>
</table>

Fig. 3. Draft of Powell’s strategy

Powell’s method belongs to the direct search methods, i.e. no first or second order derivatives are required. Here, we only state the basic idea of Powell’s method, i.e. line search along the coordinate axes in the first step and along estimated conjugate gradient directions in the following steps. Let \( x_0 \) be the initial guess of a minimum of function \( f \). At first, Powell follows successively each standard base vector until
a minimum of \( f \) is found. Hence, \( f \) becomes a one-variable function along each base vector and performs line search to find the minimum. Let \( x_0 \) be the initial candidate solution. Let \( \Pi = \{u_1, \ldots, u_N\} \) be a set of vectors that are initialized with the standard base vectors. The optimizer generates a sequence of points \( p_0, \ldots, p_N \). Figure 3 shows the pseudocode of the principle of Powell’s method. The method itself makes use of further concepts that are left out in the pseudo-code to improve readability. E.g. It discards the vector \( u_m \) with the largest decrease in \( f \) over all direction vectors in line 6. For a detailed introduction to the strategy of Powell we refer to the depiction by Schwefel [17].

### B. The ILS-Powell-Hybrid

The ILS-Powell-Hybrid proposed in this paper is based on three key concepts, each focusing on typical problems that occur in real-valued solution spaces:

- Powell’s optimization method: Powell’s method is a fast direct search optimization method – in particular appropriate for unimodal and convex fitness landscapes.
- Iterative Local Search: In order to prevent Powell’s method from getting stuck in local optima, the ILS approach starts from perturbated local solutions. Hence, ILS performs a global control of \( \lambda \) local Powell optimizers.
- Adaptive control of mutation strengths: 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.

In the previous Paragraphs we introduced the strategy of Powell and the ILS principle. Figure 4 shows the pseudo-code of the ILS-Powell hybrid. At the beginning an initial solution \( \mu \) is produced and optimized with the strategy of Powell. In an iterative loop \( \lambda \) offspring solutions \( s' \) are produced by means of Gaussian mutation with the global mutation strength \( \sigma \), i.e. each component \( x_i \in \mathbb{R} \) of \( s \) is mutated independently

\[
x'_i = x_i + \sigma \cdot \mathcal{N}(0, 1)
\]

Afterwards, \( s' \) is locally optimized with the strategy of Powell. After \( \lambda \) solutions have been produced in this kind of way, the \( \mu \) best are selected and the arithmetic mean \( \langle s \rangle \) is computed. If the search stagnates, i.e. the condition

\[
|\hat{s} - \hat{s}_{t-1}| < \theta
\]

becomes true, the mutation strength is increased by multiplication with \( \tau > 1 \)

\[
\sigma = \sigma \cdot \tau.
\]

Otherwise, the mutation strength \( \sigma \) is decreased by multiplication with \( 1/\tau \). The effect of an increase of mutation strength \( \sigma \) is that local optima can be left. A decrease of mutation strength lets the algorithm converge to the local optimum in the vicinity defined by \( \sigma \). At first, this technique seems to be in contraposition to the success rule of Rechenberg [15]. The latter decreases the step sizes in case of failure and increases the mutation strengths in case of success. This strategy is reasonable for local approximation: Smaller changes to solutions will increase the probability to be successful during approximation of local optima. But in our approach the strategy of Powell performs the approximation of the local optimum. The step control of the ILS part has another task: leaving local optima when the search stagnates. Of course, the local optimum may be the global one, but if this is the case, the technique will find the latter again and again.

### IV. Experimental Analysis

This Section provides an experimental analysis of the ILS-Powell-Hybrid, in particular in comparison to the strategy of Powell and to a standard evolution strategy, the \( (\mu, \lambda) \)-ES [3]. The experimental analysis concentrates on typical test problems known in literature. Table I shows the test problems we refer to in our analysis. Furthermore, it provides the experimental conditions, i.e. the starting point \( \hat{y}_{\text{init}} \), the initial step sizes \( \sigma_{\text{init}} \), and the termination condition \( f_{\text{stop}} \).
A. Comparison with Other Approaches

Table II shows the experimental results of the strategy of Powell, a standard \((\mu, \lambda)\)-ES and our ILS-Powell-Hybrid. The table shows the numbers of fitness functions evaluations (ffe) until the optimum is reached with accuracy \(f_{\text{stop}}\). Each technique has been run 25 times and the best, the mean value and the corresponding standard deviation is shown. For the control of \(\sigma\), see equation 3 we use the setting \(\tau = 2\). A further discussion of \(\gamma\) provides Paragraph IV-B. For the ILS-Hybrid and the ES we set \(\mu = 2\) and \(\lambda = 10\).

On the Sphere model, both variants show the fast capabilities of Powell’s search in unimodal fitness landscapes. The \((\mu, \lambda)\)-ES is able to approximate the optimum with arbitrary accuracy, but with significantly slower speed of convergence. The same can be observed on the Doublesum problem. Here, the \((\mu, \lambda)\)-ES is even slower while the Powell-based algorithms are as fast as on the Sphere model. On Rosenbrock, the \((\mu, \lambda)\)-ES turns out to be very slow while both Powell techniques show a very fast approximation behavior again. Nevertheless, the strategy of Powell is not able to find the optimum in every run, but only in 19 runs of the 25 ones\(^1\). But the strategy of Powell totally fails on the problem Rastrigin, where it does not find the optimal solution in a single run. Rastrigin exhibits many local optima where Powell’s method gets stuck into. With proper mutation parameter settings (\(\tau = 2\)), the \((\mu, \lambda)\)-ES is capable of finding the optimal solution. Lower settings for \(\tau\) lead to much slower convergence behaviors. The ILS-Powell-Hybrid is also able to find the optimum in every run, but is slightly slower – in comparison to the optimized settings of the ES. The problem of Griewank is another good example for the fast capabilities of Powell’s method. The ES is clearly outperformed by both Powell-variants. While the evolution strategy and Powell’s method completely fail to find the optimal solution of Schwefel’s function, the hybrid algorithm is able to find the optimal solution in every run.

The outcome of the experiments can be summarized as follows:

- As expected, the strategy of Powell outperforms the \((\mu, \lambda)\)-ES on unimodal functions like Sphere and Doublesum – so does the ILS-Powell hybrid.
- On highly multimodal functions like Schwefel or Rastrigin the strategy of Powell gets stuck in local optima. But the ILS-approach is able to leave these local optima and approximate the optimal solution.

In comparison to the results for the Covariance Matrix Adaptation (CMA-ES) [12] and the variant CMSA-ES – that are considered to be the state-of-the-art methods of evolutionary optimization - recently reported by Beyer [2], the ILS-Powell-Hybrid turns out to be much faster. A detailed experimental comparison will be subject to future work.

We concentrate on the behavior of the ILS-Powell-Hybrid on the highly multimodal fitness landscape of Schwefel’s function. As already reported the hybrid technique is able to leave local optima. In the upper part of figure 6 we can see the fitness development of a single run on the function Rastrigin. The strategy of Powell moves the candidate solutions into local optima. The fitness development reveals that the local optima are left and a better local optimum is found. In the lower part of figure 6 we can see the corresponding development of mutation strength \(\sigma\). When the search gets stuck in a local optimum, the strategy increases \(\sigma\) until the local optimum is successfully left and a better local optimum is found. The approach moves from one local optimum to another controlling \(\sigma\) – until the global optimum is found.

B. Strength of the Perturbation Mechanism

As pointed out in Section III the strength of the perturbation mechanism plays an essential role for the ILS mechanism. What is the influence of parameter \(\tau\) – the increase of \(\sigma\) in case of stagnation and decrease in case of an advance? We tested various settings for \(\tau\). The results of this analysis are presented in table III. The figures show the number of fitness functions evaluations until the optimal solution of the problem Rastrigin with \(N = 30\) is found. It turns out that \(\tau = 2\) is a reasonable setting. Too fast increase in case of stagnation – e.g. with \(\tau = 10\) – deteriorates the results and lets the ILS method work like a simple restart approach.

---

\(^1\)Consequently, the better mean of the strategy of Powell is not highlighted.
TABLE II
Experimental comparison of the ILS-Powell Hybrid to the strategy of Powell and a \((\mu, \lambda)\)-ES. Best, mean and dev show the number of iterations until the termination criterion is met, # states the number of runs in which the optimum is reached.

<table>
<thead>
<tr>
<th>Function</th>
<th>((\mu, \lambda))-ES</th>
<th>Powell</th>
<th>ILS-Powell-Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>best</td>
<td>mean</td>
<td>dev</td>
</tr>
<tr>
<td>Sphere</td>
<td>2.875</td>
<td>3.326</td>
<td>242.9</td>
</tr>
<tr>
<td>Doublesum</td>
<td>31,841</td>
<td>41,038.6</td>
<td>4,374.3</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>(&gt; 10^8)</td>
<td>(&gt; 10^9)</td>
<td>(&gt; 10^4)</td>
</tr>
<tr>
<td>Rastrigin</td>
<td>56,587</td>
<td>59,933.9</td>
<td>1,769.2</td>
</tr>
<tr>
<td>Griewank</td>
<td>54,411</td>
<td>60,777.3</td>
<td>2,968.0</td>
</tr>
<tr>
<td>Schwefel</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

![Fig. 6](image)

Fig. 6. Development of fitness (upper part) and step size \(\sigma\) (lower part) on the highly multimodal function Schwefel with \(N = 10\). When the search gets stuck in local optima, the perturbation mechanism increases \(\sigma\) and enables to escape from the basins of attraction.

V. Conclusion

Iterated Local Search is a successful hybridization technique in combinatorial solution spaces. In this paper we have shown that this assumption also holds true for real-valued search domains. We proposed to combine the strategy of Powell and elements from stochastic search in an ILS framework. It is worth to mention that the approach significantly outperforms the standard \((\mu, \lambda)\)-ES, and shows approximation behaviors that are superior to the Covariance Matrix Adaptation Evolution Strategy. The strategy of Powell is the reason for the power of the the ILS hybrid. Nevertheless, whenever Powell gets stuck in multimodal fitness landscapes, the adaptive perturbation mechanism helps to move out. It seems worth to try further ILS hybridizations with direct search methods such as Nelder-Mead or other techniques. A hybridization with Covariance Matrix optimizers is no reasonable undertaking as the local search method disturbs the Gaussian based update of the covariance matrix – and our experimental analysis confirmed that no further improvement can be gathered in comparison to the approach at hand.

REFERENCES