

Winker, Peter

Working Paper

The Stochastics of Threshold Accepting: Analysis of an Application to the Uniform Design Problem

Discussion Paper, No. 2005,003E

Provided in Cooperation with:

University of Erfurt, Faculty of Economics, Law and Social Sciences

Suggested Citation: Winker, Peter (2005) : The Stochastics of Threshold Accepting: Analysis of an Application to the Uniform Design Problem, Discussion Paper, No. 2005,003E, Universität Erfurt, Staatswissenschaftliche Fakultät, Erfurt

This Version is available at:

<https://hdl.handle.net/10419/23940>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



UNIVERSITÄT
ERFURT

Staatswissenschaftliche Fakultät

Faculty of Economics,
Law and Social Sciences

Discussion Paper No.: 2005-003E

**The Stochastics of Threshold
Accepting: Analysis of an
Application to the Uniform Design
Problem**

Peter Winker

Für den Inhalt des Diskussionspapiers sind die jeweiligen Autoren/innen allein verantwortlich.

ISSN 1610-9198 (Print)

ISSN 1610-918X (Internet)

The Stochastics of Threshold Accepting: Analysis of an Application to the Uniform Design Problem*

Peter Winker[†]

March 2, 2005

Abstract

Threshold Accepting (TA) is a powerful optimization heuristic from the class of stochastic local search algorithms. It has been applied successfully to different optimization problems in statistics and econometrics, including the uniform design problem. Using the latter application as example, the stochastic properties of a TA implementation are analyzed. We provide a formal framework for the analysis of optimization heuristics like TA, which can be used to estimate lower bounds and to derive convergence results. It is also helpful for tuning real applications. Based on this framework, empirical results are presented for the uniform design problem. In particular, for two problem instances, the rate of convergence of the algorithm is estimated to be of the order of a power of -0.3 to -0.7 of the number of iterations.

Keywords: Heuristic optimization; Threshold Accepting; Stochastic analysis of heuristics.

*I am indebted to M. Gilli and D. Maringer for helpful comments on a preliminary version of this paper.

[†]Department of Economics, Law and Social Sciences, University of Erfurt

1 Introduction

Threshold Accepting (TA) is an optimization heuristic.¹ It might be considered as a modification of the more often used Simulated Annealing (SA) heuristic (Kirkpatrick *et al.*, 1983) using a deterministic acceptance criterion instead of the probabilistic one in SA. Nevertheless, it shares the property of most optimization heuristics to provide a stochastic approximation to the optimum, i.e. results will change when the algorithm is run repeatedly due to random effects in the start point and search steps.

Given that classical or standard optimization techniques such as Newton's method and its descendants fail to solve many of the problems arising in statistics and economics, it seems adequate to also consider optimization heuristics. For example, TA has been successfully applied to many problems in statistics and econometrics, which could not be tackled by standard optimization algorithms (Winker, 2001). An overview on applications in statistics and econometrics as well as a detailed description of the algorithm and its implementation are provided by Winker and Maringer (2005).

Optimization heuristics like TA are used whenever conventional optimization tools fail to provide optimal solutions. Optimization heuristics are appealing because they can be applied in such situations, including the case when theoretical complexity analysis indicates that an exact optimum solution cannot be obtained by deterministic methods in reasonable time (NP-complete problems).

As an example, we consider the application of TA in the context of constructing low discrepancy experimental designs. TA has been applied by Winker and Fang (1997*a*) to obtain lower bounds for the star-discrepancy, while Winker and Fang (1997*b*) use the approach to obtain low discrepancy U -type designs for the star-discrepancy. Fang *et al.* (2000) extended the analysis to several modifications of the L_2 -discrepancy, while Fang *et al.* (2002) and Fang *et al.* (2003) consider the centered and wrap-around L_2 -discrepancy. Here, we use settings from Fang *et al.* (2003) for the empirical demonstration of the formal framework. This example differs from the standard scenario due to the fact, that the analysis by Fang *et al.* (2003) provides theoretical lower bounds for the objective function to be minimized. Consequently, a result obtained by the optimization heuristic, which meets this

¹For an introduction to optimization heuristics and their classification see Winker and Gilli (2004).

lower bound, represents a proven local minimum. However, given that the lower bounds cannot be met for all problem instances, the inverse does not hold true.

The paper is structured as follows. Section 2 introduces a formal framework for the analysis of stochastic optimization algorithms. Section 3 provides results for an implementation to the uniform design problem. An outlook to further research is provided in Section 4 together with some concluding remarks.

2 Formal Framework

Let us assume that TA is implemented for the purpose of minimizing an objective function f over a search space Ω , which can be finite, discrete or continuous. Of course, by replacing f with $-f$, the algorithm can also be applied to maximization problems. Let f_{min} denote the minimum of f over the search space if existent, otherwise the infimum. A single run of an optimization heuristic like TA will provide an approximation ζ to this minimum.

Of course, the quality of this approximation will depend on implementation details, which are typically covered to some extent in the description of the algorithm.² Furthermore, in the case of a local search heuristic like TA, the quality of the approximation will also depend on the number of iterations I , which, in general, is almost proportional to the computational time. Thus, let ζ_I denote an approximation to the minimum obtained by a TA implementation with I iterations.

Optimization heuristics comprise random elements. First, they start with a random candidate solution. Second, the generation of new candidate solution during a local search typically also includes random elements.³ Finally, while the acceptance criterion is deterministic for TA, it might comprise further randomness for other algorithms, as does, e.g., SA. Thus, for a given implementation of TA with all parameters set and a given number of iterations, the result of a single run ζ_I should be considered as a realization of a random variable.

²A detailed description for the TA heuristics can be found in Winker (2001).

³An exception is the variant of TA discussed by Dueck and Scheuer (1990) with a purely deterministic selection of new candidate solutions.

Taking into account this stochastic property,⁴ the application of TA to an optimization problem should be considered as a stochastic mapping

$$TA : \Omega \rightarrow \zeta_I, \zeta_I \sim D_I^{TA}(\mu_I, \sigma_I) \quad (1)$$

mapping the search space Ω to a random realization ζ_I depending on the random number sequence used for the specific run. Thereby, ζ_I is a random realization from a distribution $D_I^{TA}(\mu_I, \sigma_I)$. For ease of notation, we will omit the superscript TA in the rest of the paper. Of course, this distribution is truncated from the left at the value of the global minimum $f_{min} = \inf\{f(x)|x \in \Omega\}$. If Ω is finite, the density of D_I at f_{min} is positive for any I . The same holds true for continuous sets Ω if f satisfies standard regularity conditions.

Although this framework applies to any application of optimization heuristics, in general, it is not mentioned. Instead, reported results are often restricted to a single estimate ζ_{min} , which is the minimum obtained over a finite number of runs, R , of the algorithm with a given number of iterations I , i.e. $\zeta_{min}^R = \min\{\zeta_I^i | i = 1, \dots, R\}$. Sometimes, R is also reported. In the framework introduced above, we would consider the set $\{\zeta_I^i | i = 1, \dots, R\}$ as a random sample from $D_I(\mu_I, \sigma_I)$. Given that this distribution is left truncated, the minimum over the set, ζ_{min}^R , in general, represents the maximum likelihood estimator of f_{min} under the assumption of a given left truncated distribution. It can also be interpreted as the first order statistic. Nevertheless, one should be interested in the properties of this estimator except for cases when it is obvious that $\zeta_{min}^R = f_{min}$. Such a situation can be given if a theoretical lower bound is available for f_{min} as in the applications to the uniform design problems discussed by Fang *et al.* (2003) and Fang *et al.* (2005). However, for most applications in statistics and econometrics, no binding lower bounds are known. Thus, a more detailed analysis of the properties of ζ_{min}^R is required. In particular, the rate of convergence of ζ_{min}^R to f_{min} as $I \rightarrow \infty$ is of crucial interest. Furthermore, it might be interesting to study the properties of alternative estimators of f_{min} .

Before turning to an empirical analysis of $D_I(\mu_I, \sigma_I)$ in Section 3 for the application to the uniform design problem, a few general properties of this distribution can be discussed. For TA, Althöfer and Koschnik (1991) provide

⁴The same reasoning also applies to implementations of classical optimization algorithms, if the results depend on the initialization of the algorithm, which might be chosen at random.

a convergence result. The theorem states that there exist suitable parameters for the TA implementation such that for any $\varepsilon > 0$ and $\delta > 0$, there exists a number of iterations I such that

$$Prob(|\zeta_I - f_{min}| < \varepsilon) > 1 - \delta.$$

Consequently, when $I \rightarrow \infty$, all quantiles of $D_I(\mu_I, \sigma_I)$ converge to f_{min} , while, in general, it is not possible to make a statement about μ_I or σ_I . If, however, the search space Ω is finite, and, consequently, the support of $D_I(\mu_I, \sigma_I)$ has to be finite, too, μ_I will converge to f_{min} and σ_I will converge to zero as $I \rightarrow \infty$.

In real applications, one is not interested in obtaining a good estimate of μ_I , but a good upper bound for f_{min} . Therefore, the standard proceeding in applications of optimization heuristics might be justified, namely to sample several drawings from $D_I(\mu_I, \sigma_I)$ and to use the minimum of this sample as estimate for f_{min} . In fact, while the expectation of a single drawing corresponds to an estimate of μ_I , the minimum of a sample of several replications, i.e. the first order statistic from a sample of R replications, corresponds to an estimate of some lower quantile of $D_I(\mu_I, \sigma_I)$, which converges to f_{min} with the number of iterations I growing to infinity. Even if μ_I also converges to f_{min} , e.g. in the case of a finite Ω , the lower quantiles will converge faster.

Thus, it remains an important question to be answered how to distribute some given amount of computational resources, say C (measured in iterations), on replications R and iterations per replication I . If all resources are spent on a single run, this run provides an approximation to μ_C . If the resources are split on R replications, each run provides an approximation to $\mu_{C/R}$ and all replications together provide an empirical estimate of $D_{C/R}(\mu_{C/R}, \sigma_{C/R})$ which allows to obtain an estimate of the $1/R$ quantile.

Winker (2001, pp. 129ff) considers an application to a traveling salesman problem and finds that from a set of possible choice for R of $\{1, 10, 100\}$, 10 appeared to be the optimal choice. However, this results cannot be easily generalized to different problem instances. In the next section, a different application, the uniform design problem, is considered. Parametric and non-parametric approximations to D_I will be used to derive an optimal tradeoff for I and R for this application. Furthermore, the speed of convergence of the parameters of D_I will be analyzed by means of linear regression.

3 Results for Uniform Design Implementation

The empirical results in this section are based on an application of TA to the uniform design problem described in Fang *et al.* (2003). The task consists in distributing n points in a d -dimensional unit cube as “uniformly” as possible. Consequently, the objective function is a measure of discrepancy between the empirical distribution of the given point set and the theoretical uniform distribution on the unit cube. For specific measures of this discrepancy in the L_2 -norm, Fang *et al.* (2003) and Fang *et al.* (2005) provide theoretical lower bounds for the objective function for given values of d and n . Hence, it is possible to express the value of the objective function obtained by a run of TA relative to this lower bound. If the lower bound is met, the corresponding result is a global optimum. Unfortunately, it cannot be guaranteed that for any d and n , there exists a solution meeting the theoretical lower bound.

3.1 Empirical Distribution

Here, two instances of the uniform design problem with three levels for the wrap-around L_2 -discrepancy as objective function are considered. For the first instance, the size of the experimental design is $d = 6$ factors and $n = 12$ runs. For this instance, the lower bound can be obtained for some runs of the algorithm already for small numbers of iterations. In fact, for the present implementation, this lower bound could be reached already in one of 1 000 runs with only 500 iterations. However, as pointed out in the previous section, looking at this empirical minimum of several runs does not provide all relevant information and might not be robust. In fact, the lower bound could not be reached in a single run out of 1 000 runs with 1 000 iterations, while it has been obtained repeatedly for runs with 5 000 iterations and more. Descriptive statistics for 1 000 replications for different values of I are reported in the left part of Table 1, while the right part reports results for the second problem instance with $d = 10$ and $n = 15$. A more substantial information is provided by the empirical distribution functions for ζ_I for differing I , which are shown in Figure 1.

From the plot it becomes obvious that as I increases, the distribution ζ_I approaches a degenerate distribution at $f_{min} = 0$. This supports the theoretical convergence result presented in the previous section. For a slightly larger

Table 1: Minimum, mean, and standard deviations

Iterations I	$d = 6, n = 12$			$d = 10, n = 15$		
	Minimum	Mean	Std.dev.	Minimum	Mean	Std.dev.
200	0.118	2.469	0.783	3.925	6.531	0.778
500	0.000	1.501	0.611	3.335	5.324	0.628
1 000	0.059	0.971	0.469	3.067	4.646	0.539
5 000	0.000	0.244	0.244	2.400	3.406	0.353
10 000	0.000	0.217	0.193	2.263	3.066	0.290
50 000	0.000	0.076	0.057	1.881	2.574	0.235
100 000	0.000	0.060	0.036	1.861	2.439	0.201
500 000	0.000	0.033	0.032	1.345	2.246	0.194
1 000 000	0.000	0.020	0.028	1.613	2.171	0.182
5 000 000	n.a.	n.a.	n.a.	1.345	2.020	0.172

problem instance, i.e. $d = 10$ and $n = 15$, a similar plot of the empirical distributions of relative deviations from a theoretical lower bound is shown in Figure 2. Again, it can be observed how the empirical distributions shift left and become steeper, i.e. μ_I and σ_I decrease, as I increases. However, even for $I=5\,000\,000$, the theoretical lower bound is never met, and the empirical distribution does not exhibit a clear cut minimum, which is reached repeatedly. Thus, although it cannot be guaranteed that the theoretical lower bound can be actually met, one might expect that a further increase in I would result in a further shift of ζ_I .

3.2 Approximation by Truncated Normal Distributions

The next step of the analysis consists in fitting parametric distributions to the empirical distributions obtained for different values of I . Obviously, only left-truncated distributions should be considered. Furthermore, given that the search space Ω is finite for both instances of the uniform design problem, the set of possible values of the objective function is finite, too. Consequently, the distribution functions ζ_I cannot be continuous, but exhibit discrete steps at each possible value of the objective function. Nevertheless, given the huge cardinality of Ω and the resulting small distances between steps in

Figure 1: Empirical distribution of ζ_I for different I ($d=6, n=12$)

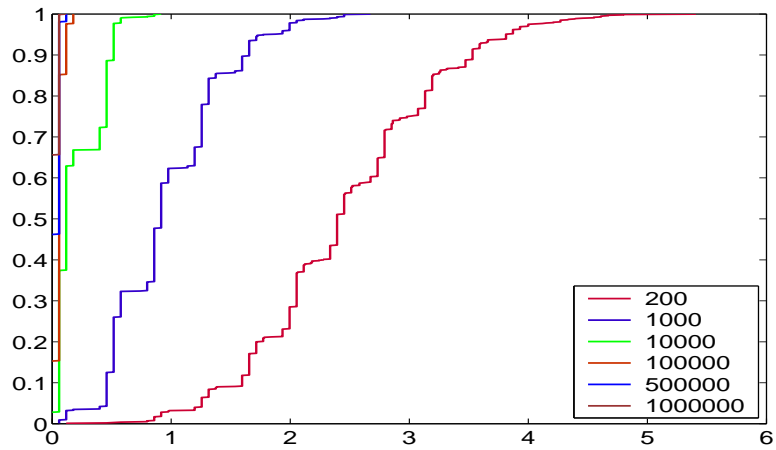
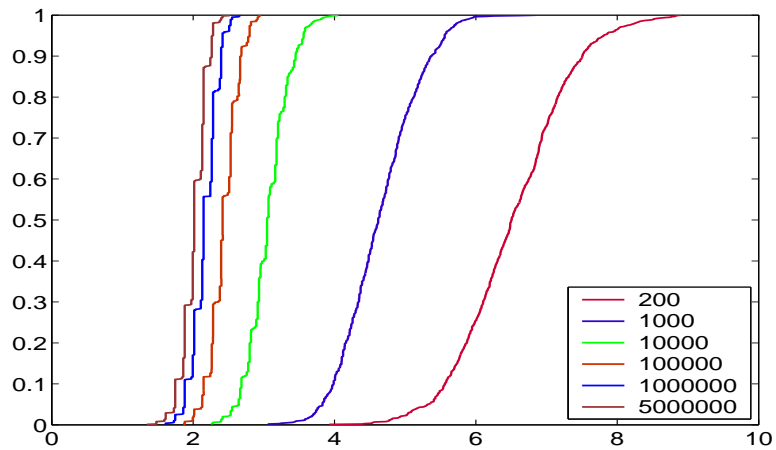


Figure 2: Empirical distribution of ζ_I for different I ($d=10, n=15$)



the distribution function ζ_I , an approximation by a continuous distribution function appears admissible.

Thus, in a first step, for each empirical distribution function ζ_I of both problem instances, the parameters of a truncated normal distribution are fitted by a maximum likelihood estimation. Obviously, the maximum likelihood estimator of the truncation value corresponds to the empirical minimum of the observed values (Schneider, 1986, p. 47). For the first problem instance with $d = 6$ and $n = 12$, the theoretical lower bound of zero is met already by a run with only 500 iterations. If the lower bound is met, it does not represent an estimator of the truncation value, but a truncation value of zero can be considered as being given. For the second instance with $d = 10$ and $n = 15$, the theoretical lower bound is never met. Since theory does not guarantee that there exists a solution meeting the lower bound, the truncation value remains unknown and has to be estimated. For ease of comparison, we assume the truncation parameter as unknown for both cases. Furthermore, the estimation is based only on the empirical results for given I , i.e. the additional knowledge about lower bounds obtained by runs with a different number of iterations is not taken into account.⁵

Table 2 summarizes the estimation results for the first problem instance ($d = 6, n = 12$), while the corresponding values for $d = 10$ and $n = 15$ are reported in Table 3. Thereby, “Truncation”, μ_I , and σ_I are the estimated parameters of the truncated normal distribution, \bar{X} and $s(X)$ the empirical moments and $E[X]$ and $\sqrt{V[X]}$ the moments of the truncated normal distribution.

In both cases, the truncated normal distribution provides a reasonable approximation to D_I . Since the truncation often becomes binding for the first problem instance already for small values of I , the expectation of the truncated distribution differs considerably from the expectation of the underlying normal distribution (μ_I). This effect is much less pronounced for the second example, where the estimated density at the truncation values is small for all I .

Given that D_I is defined on a finite support for the specific application, it is not too surprising that a standard goodness-of-fit test rejects the hypothesis that the empirical data are generated from a truncated normal distribution except for the case D_{200} for the second problem instance, when the

⁵The effect of considering this additional information will be considered when analyzing the distribution of order statistics below.

Table 2: Truncated normal distributions for $d = 6, n = 12$

Iterations I	Truncation	μ_I	\bar{X}	$E[X]$	σ_I	$s(X)$	$\sqrt{V[X]}$
200	0.118	2.465	2.469	2.470	0.789	0.783	0.784
500	0.000	1.486	1.501	1.511	0.630	0.611	0.611
1 000	0.059	0.914	0.971	1.024	0.522	0.469	0.469
5 000	0.000	0.115	0.244	0.726	0.369	0.244	0.244
10 000	0.000	0.000	0.217	0.798	0.290	0.193	0.175
50 000	0.000	0.012	0.076	0.725	0.090	0.057	0.057
100 000	0.000	0.049	0.060	0.298	0.044	0.036	0.036
500 000	0.000	0.000	0.033	0.798	0.046	0.032	0.027
1 000 000	0.000	0.000	0.020	0.798	0.035	0.028	0.021

truncation is not really binding. The discrepancy is particularly strong for the lower quantiles. Although the lower quantiles of the estimated truncated normal distributions tend to be slightly smaller than the empirical quantiles as long as the empirical quantiles have positive values, a regression of empirical quantiles on the quantiles predicted by the truncated normal distribution does not find a significant bias while exhibiting a very high share of explained variance ($R^2 = 0.996$ and 0.998 for the 1%-quantile and the two problem instances, respectively).

Thus, we continue to use this approximation in the following subsections besides the empirical distribution function. Obviously, for practical applications, estimates of the truncated normal distribution can be obtained at lower computational cost than the complete empirical distribution for a considerable number of replications.

3.3 Distribution of Order Statistics

Returning to the question on how to distribute some given amount of computational resources on replications R and iterations per replication I , the distribution of lower quantiles or – empirically – order statistics is of specific relevance. Therefore, in this subsection, results on the distribution of order statistics are presented both based on the empirical distribution function and on the estimated truncated normal distribution function.

Table 3: Truncated normal distributions for $d = 10, n = 15$

Iterations I	Truncation	μ_I	\bar{X}	$E[X]$	σ_I	$s(X)$	$\sqrt{V[X]}$
200	3.925	6.530	6.531	6.531	0.780	0.778	0.778
500	3.335	5.323	5.324	5.325	0.630	0.628	0.627
1 000	3.067	4.643	4.646	4.649	0.543	0.539	0.538
5 000	2.400	3.403	3.406	3.411	0.357	0.353	0.353
10 000	2.263	3.063	3.066	3.073	0.294	0.290	0.290
50 000	1.881	2.573	2.574	2.578	0.237	0.235	0.235
100 000	1.861	2.437	2.439	2.444	0.203	0.201	0.201
500 000	1.345	2.246	2.246	2.246	0.194	0.194	0.194
1 000 000	1.613	2.170	2.171	2.174	0.183	0.182	0.182
5 000 000	1.345	2.020	2.020	2.020	0.172	0.172	0.172

If R replications are performed for a given number of iterations I , we obtain a random sample ζ^1, \dots, ζ^R from D_I . Let $\zeta^{(1)}, \dots, \zeta^{(R)}$ denote the ordered values. This allows to estimate a lower quantile of D_I by the first order statistic $\zeta^{(1)}$. The marginal probability density function of $\zeta^{(1)}$ is given by

$$f_1(x) = d_I(x)(1 - D_I(x))^{R-1}R, \quad (2)$$

where $d_I(x)$ denotes the density function of the distribution D_I .⁶

Tables 4 and 5 provide the mode and expected value of the first order statistics for the two problem instances based on the approximation by a truncated normal distribution introduced in the previous subsection. Alternatively, $D_I(x)$ can be approximated by the empirical cumulative density function and $d_I(x)$ by some kernel density estimator. However, the estimation of the expected value based on this approximation is complicated by the missing smoothness of the empirical cumulative density function. Thus, in both tables, only an estimate of the mode based on this non parametric approach is added.

⁶For a general k -th order statistic, the corresponding density is given by

$$f_k(x) = D_I(x)^{k-1}d_I(x)(1 - D_I(x))^{R-k} \frac{R!}{(k-1)!(R-k)!}.$$

Table 4: Distribution of first order statistics ($d = 6, n = 12$)

Iterations I	$R = 10$			$R = 1000$		
	trunc. normal		kernel	trunc. normal		kernel
	mean	mode	mode	mean	mode	mode
200	1.253	1.272	1.593	0.118	0.232	0.455
500	0.591	0.587	0.798	0.000	0.000	0.057
1 000	0.214	0.312	0.457	0.059	0.101	0.116
5 000	0.000	0.000	0.058	0.000	0.000	0.058
10 000	0.000	0.000	0.058	0.000	0.000	0.058
50 000	0.000	0.000	0.058	0.000	0.000	0.058
100 000	0.000	0.000	0.058	0.000	0.000	0.058
500 000	0.000	0.000	0.041	0.000	0.000	0.041
1 000 000	0.000	0.000	0.000	0.000	0.000	0.000

Table 5: Distribution of first order statistics ($d = 10, n = 15$)

Iterations I	$R = 10$			$R = 1000$		
	trunc. normal		kernel	trunc. normal		kernel
	mean	mode	mode	mean	mode	mode
200	5.444	5.336	5.400	4.115	4.159	4.389
500	5.001	4.363	4.500	3.361	3.467	3.674
1 000	3.934	3.824	3.828	3.067	3.134	3.220
5 000	2.875	2.868	2.892	2.400	2.436	2.414
10 000	2.747	2.626	2.645	2.263	2.287	2.282
50 000	2.388	2.215	2.259	1.881	1.911	2.014
100 000	2.115	2.132	2.128	1.861	1.883	1.881
500 000	1.970	1.946	1.994	1.640	1.616	1.727
1 000 000	1.875	1.892	1.860	1.613	1.643	1.707
5 000 000	1.776	1.756	1.747	1.480	1.470	1.478

For the smaller problem instance, the results in table 4 exhibit slight differences between the different estimates of mode and mean based on parametric and nonparametric approximations. However, these differences can mainly be attributed to border effects for the normal kernel used in the nonparametric approach when a relevant share of observations is at or close to the lower bound zero. For the larger problem instance, when the theoretical lower bound is never reached even for a large number of iterations I , this effect is less pronounced. Consequently, the results appear more robust in table 5. Furthermore, for the second problem instance, this finding allows for the conclusion that the distributions of order statistics are almost symmetric. For both problem instances and the two distinct numbers of replications considered ($R = 10$ and $R = 1000$), in general, the expected first order statistic decreases with an increase of the number of iterations per replication.

However, for the practical trade-off between replications R and number of iterations per replication I for a given total amount of computational resources C , this finding is not relevant. For this purpose, we consider the situation of a fixed amount of computational resources $C = 20\,000\,000$ and calculate the expected first order statistic for different pairs (R, I) satisfying the constraint $R \times I = C$. Table 6 summarizes the results.

Table 6: Expected first order statistics for given $C = 20\,000\,000$

R	I	$d = 6, n = 12$			$d = 10, n = 15$		
		trunc. normal		kernel	trunc. normal		kernel
		mean	mode	mode	mean	mode	mode
50 000	200	0.118	0.121	0.455	3.925	3.935	4.389
20 000	500	0.000	0.000	0.057	3.335	3.346	3.674
10 000	1 000	0.059	0.060	0.116	3.067	3.076	3.220
2 000	5 000	0.000	0.000	0.058	2.400	2.420	2.414
1 000	10 000	0.000	0.000	0.058	2.263	2.287	2.282
200	50 000	0.000	0.000	0.058	2.008	1.971	2.014
100	100 000	0.000	0.000	0.058	1.988	1.961	1.994
20	500 000	0.000	0.000	0.041	1.907	1.883	1.860
10	1 000 000	0.000	0.000	0.000	1.875	1.891	1.860
2	5 000 000	n.a.	n.a.	n.a.	2.427	1.923	1.995

For the smaller problem instance ($d = 6, n = 12$), an optimal solution can be found with high probability already for a low number of iterations. Thus, for a large amount of computational resources $C = 10\,000\,000$, the distribution on replications R and iterations I has not a major impact on the distribution of first order statistics. As long as a minimum number of 5 000 iterations is chosen, one might expect to find the true global optimum in at least one of $R = C/I$ replications.

The situation is different for the larger problem instance, where the theoretical lower bound could not be achieved even when using $I = 5\,000\,000$. Furthermore, the distribution of results does not exhibit a pronounced clustering at some lower values. For this case, the results in table 6 indicate that it is neither optimal to perform a large number of replications with a small number of iterations, i.e. sampling from a distribution D_I with high mean and variance, nor to concentrate the available resources on few replications. Rather, the optimum number of replications seems to be of the order of 10 to 20. This result provides a strong support for the typical application of local search heuristics, i.e. running the algorithm several times and reporting the best result of all runs.

3.4 Convergence

The data presented in Table 1 can also be used to estimate the rate of convergence of μ_I and σ_I . For this purpose, nonlinear least squares estimation of the model

$$\mu_I = \beta_0 + \beta_1 I^{-\beta_2} + \varepsilon_I \quad (3)$$

have been performed. For the first problem instance, the restriction $\beta_0 = 0$ has been imposed, because for this instance the lower bound can be reached. The rate of convergence β_2 is estimated to be 0.59 for this instance. Relaxing the restriction $\beta_0 = 0$ results in an insignificant estimate of β_0 , while the estimator of β_2 becomes 0.58. The R^2 of the model is above 0.999 both with and without the restriction imposed. Thus, the estimated rate of convergence for the first problem instance is faster than the standard rate of $1/\sqrt{I}$.

For the second problem instance, the true lower bound is not known. The best value found during the experiments corresponds to a 1.345% deviation from the theoretical lower bound. Consequently, model (3) is estimated both with unconstrained β_0 and $\beta_0 = 1.345$. The estimation with unconstrained β_0 results in estimators $\hat{\beta}_0 = 1.891$, i.e. above the best known solution, and

$\hat{\beta}_2 = 0.343$. Imposing the restriction $\beta_0 = 1.345$ reduces the estimated rate of convergence to $\hat{\beta}_2 = 0.250$ while the model's R^2 shrinks from 0.999 to 0.988. Finally, imposing the restriction $\beta_2 = 0$, i.e. assuming that the theoretical lower bound can be achieved, reduces the R^2 further to 0.945, while the estimated rate of convergence shrinks to 0.140. Using a likelihood ratio test, both restrictions have to be rejected. Obviously, further results for higher values of I are required to obtain a clear cut result for this more complex instance.

For practical applications, the convergence of the mean might be of less interest than the convergence of some lower quantiles. Therefore, the convergence analysis has been repeated with the empirical 1%-, 5%- and 10%-quantiles as dependent variables in equation (3). Again, for the first problem instance, the restriction $\beta_0 = 0$ is imposed, while β_0 is estimated for the second problem instance. The estimated rate of convergence $\hat{\beta}_2$ is higher than 0.7 for the first problem instance and still faster than 0.3 for the second problem instance for all three considered quantiles.

4 Conclusions and Outlook

Optimization heuristics like TA are applied more and more often to optimization problems in statistics and econometrics which cannot be tackled by classical tools. However, the ease of implementation and versatility of these new tools comes at the cost of introducing an additional stochastic component in the analysis. In this paper, a formal framework is introduced to discuss this additional source of randomness.

Based on an example from uniform design theory, practical applications of the formal framework are demonstrated. It allows to derive optimal restarting schemes by estimating order statistics. Furthermore, the rate of convergence of the algorithm can be estimated for the given problem instances.

The first results obtained in this paper are very promising. Thus, straightforward extensions of the approach will be analyzed in the next step. First, the approximation of the empirical result distribution by a truncated normal distribution will be tested and alternative distributions will be considered. Second, for the specific application, the analysis will be repeated for the absolute values of the objective function instead of the relative deviations from the theoretical lower bound. Then, the results can be generalized more easily to other problems, for which no lower bounds are available. Finally, for other

applications of optimization heuristic, e.g., maximum likelihood estimation in GARCH–settings or estimation of censored quantile regressions, it is not the convergence of the values of the objective function to its global optimum, but the convergence of estimated parameters to their “true” values which is of interest. This convergence can also be considered in the framework introduced in this paper.

References

- Althöfer, I. and K.-U. Koschnik (1991). On the convergence of threshold accepting. *Applied Mathematics and Optimization* **24**, 183–195.
- Dueck, G. and T. Scheuer (1990). Threshold accepting: A general purpose algorithm appearing superior to simulated annealing. *Journal of Computational Physics* **90**, 161–175.
- Fang, K.-T., D. K. J. Lin, P. Winker and Y. Zhang (2000). Uniform design: Theory and application. *Technometrics* **42**, 237–248.
- Fang, K.-T., X. Lu and P. Winker (2003). Lower bounds for centered and wrap-around L_2 -discrepancies and construction of uniform designs by threshold accepting. *Journal of Complexity* **19**, 692–711.
- Fang, K.-T., C.-X. Ma and P. Winker (2002). Centered L_2 -discrepancy of random sampling and latin hypercube design, and construction of uniform designs. *Mathematics of Computation* **71**, 275–296.
- Fang, K.-T., D. Maringer, Y. Tang and P. Winker (2005). Lower bounds and stochastic optimization algorithms for uniform design with three and four level. *Mathematics of Computation* forthcoming.
- Kirkpatrick, S., C. Gelatt and M. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671–680.
- Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations*. Dekker. New York.
- Winker, P. (2001). *Optimization Heuristics in Econometrics: Applications of Threshold Accepting*. Wiley. Chichester.

- Winker, P. and K.-T. Fang (1997a). Application of threshold accepting to the evaluation of the discrepancy of a set of points. *SIAM Journal on Numerical Analysis* **34**, 2028–2042.
- Winker, P. and K.-T. Fang (1997b). Optimal U-type designs. In: *Monte Carlo and Quasi-Monte Carlo Methods 1996* (H. Niederreiter, P. Hellekalek, G. Larcher and P. Zinterhof, Eds.). Lecture Notes in Statistics 127. Springer. New York. pp. 436–448.
- Winker, P. and M. Gilli (2004). Applications of optimization heuristics to estimation and modelling problems. *Computational Statistics & Data Analysis* **47**(2), 211–223.
- Winker, P. and D. Maringer (2005). Threshold accepting in economics and statistics. In: *Proceedings CMS 2005* (E. Konthoghiorges, Ed.). forthcoming.