Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

ELSEVIER International Journal of Approximate Reasoning 35 (2004) 223–238

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Adaptive multiresolution search: How to beat brute force?

Marc Thuillard *

BELIMO Automation AG, Brunnenbachstrasse 1, CH-8340 Hinwil, Switzerland

Received 1 February 2003; accepted 1 August 2003

Abstract

Multiresolution and wavelet-based search methods are suited to problems for which acceptable solutions are in regions of high average local fitness. In this paper, two different approaches are presented. In the Markov-based approach, the sampling resolution is chosen adaptively depending on the fitness of the last sample(s). The advantage of this method, behind its simplicity, is that it allows the computation of the discovery probability of a target sample for quite large search spaces. This permits to "reverse-engineer" search-and-optimization problems. Starting from some prototypic examples of fitness functions the discovery rate can be computed as a function of the free parameters. The second approach is a wavelet-based multiresolution search using a memory to store local average values of the fitness functions. The sampling density probability is chosen per design proportional to a low-resolution approximation of the fitness function. High average fitness regions are sampled more often, and at a higher resolution, than low average fitness regions. If splines are used as scaling mother functions, a fuzzy description of the search strategy can be given within the framework of the Takagi–Sugeno model.

Keywords: Adaptive search; Wavelet; Multiresolution; Uncertainty; Fuzzy

* Tel.: +41-1938-4411; fax: +41-1938-4710.
E-mail address: marc.thuillard@belimo.ch (M. Thuillard).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2003.08.003

1. Introduction

Humans (or in a broader sense, "nature") elaborate complex strategies to gather and organize efficiently useful information based on experience, knowledge, partial information, in-born capabilities or even intuition. In search and learning problems, one of the main strategies used by nature is the so-called "trial and error". Adaptive search methods represent a tentative to apply in computation the trial and error approach. The basic idea behind an adaptive search is to extract information from the previously sampled points and to include that information, either implicitly or explicitly, in the search strategy. In recent years, much interest has been concentrated on adaptive search methods, in which the information is extracted implicitly. Many such approaches have been inspired by nature. Genetic algorithms, multiresolution search algorithms and simulated annealing are just some examples. A large number of questions in optimization and search problems reduce to the following problem: consider an hyper surface in $\Re^n$ (or in Z^n). Assume that the surface can be described by an expression of the form $y = f(x)$, with $x \in \Re^{n-1}$ (or Z^{n-1}), in which $f(x)$ is a measure of the fitness or goodness of a solution x, taking typically values between zero and one. We want to find values of x for which either f is maximal or at least within a small range of the maximal value.

Without preliminary knowledge, the optimal search method consists in avoiding testing a possible solution twice. It is only possible to "beat brute force" if some prior information on the fitness function is available. This fact has been expressed under different forms that can be regrouped under the general class of "no free lunch theorems" [1]. The mechanisms and assumptions behind adaptive search are far from being perfectly identified or understood. Probably the most important general question is to identify what kind of information permits to elaborate an efficient adaptive search strategy? Efficient search strategies can be designed if some known relationships between good solutions can be reasonably postulated. The relationships characterizing good solutions in the search space cover a very large range of different possibilities. Let us give here a few examples: good solutions are found in regions of high average fitness; the problem is separable into several of its variables; a strong correlation exists between several variables. The above statements have in common, that they are all expressions in which the knowledge of a number of sample points will permit to either exclude possibilities, to limit the range of parameters or to lead the search algorithm in regions in which good solutions are likely to be found. By doing so, the cumulative probability of sampling good solutions increases beyond the level it would have if the search had been totally random. A central question on the research agenda is therefore to determine which information characterizes a certain type of search problems and how to use that information to reduce the computing power necessary to achieve one's goal. Adaptive search methods are all based implicitly or explicitly on some assumption or model. Several variants of genetic algorithms are based on the so-called building block hypothesis [2]. Local search methods, such as hill climbing [3,6], assume the existence of a limited number of local maxima. Ants search [4] can be regarded as a local search combining projections of good solutions on low-dimension spaces (the edges of a graph). Dynamic programming [5] uses the knowledge that the search space can be reduced algorithmically or stated slightly differently that the search problem can be decomposed into a sequence of decisions. The multiresolution search methods that will be introduced below are based on a simple assumption: a proximity relation is assumed between high-fitness samples. More precisely, the assumption is that at some scale, target samples are found with a higher probability in domains of high average fitness.

Some of the most successful approaches in search-and-optimization problems use some stochastic elements [2–10]. Genetic algorithms, multiresolution search [7], gradient search, simulated annealing [9] are some examples. The performances of these algorithms depend quite significantly on the choice of some free parameters. Presently one relies much on "good practices" obtained through numerical experiments on a number of typical problems to determine these parameters. A quite disturbing fact is that even when the fitness function is known, it is generally impossible to determine even a posteriori if the chosen parameters were appropriate or even sometimes if the search is on average better than a search with brute force. As an example, let us discuss succinctly genetic algorithms. Even in the simple genetic algorithm [10], the computation of the expected performances is limited to very small problems' size. For large problems, the transition matrix describing the stochastic process has about $\text{Size}^{2\beta}$ elements ("Size" is the number of elements in the search space and β the number of elements being processed at each iteration). For instance, for a problem with a search space containing 1000 elements, from which 10 are processed at each iteration, the transition matrix of the Markov process of the simple genetic algorithm contains about 10^{60} elements. In order to better comprehend the performances of adaptive stochastic search, it is important to develop methods for which the expected probability of discovering a target sample can be computed for quite large search space from the knowledge of the fitness surface. This permits to compare the search quality on different fitness surfaces and to develop an understanding on how to choose the correct parameters.

In Section 2, Markov-based multiresolution search algorithms are presented. Besides being easy to implement the Markov-based approach possesses an important feature. Contrarily to most stochastic methods, the probability of discovering a target sample can be computed for quite large search spaces. For that reason, multiresolution search models may be used as prototypic models to study quantitatively the performance of adaptive search models.

Section 3 presents another approach to multiresolution and wavelet-based search methods using elements of wavelet estimation theory [7]. This approach represents a natural connection between search-and-optimization theory and functional data analysis. Contrarily to the Markov approach, the transient properties of the probability of discovering a target solution cannot be computed easily. The sampling probability at equilibrium can only be calculated. If splines are used as mother scaling functions a fuzzy interpretation of the results can be given within the framework of the Takagi–Sugeno model.

2. Markov-based multiresolution search

The complementarity between multiresolution analysis and adaptive search techniques has been recognized for already quite some time. Bethke [11] introduced Walsh partition functions in the field of genetic algorithms. Important insights on the building block hypothesis were gained using this approach. Problems, that are intrinsically difficult for genetic algorithms, were designed using the Walsh functions [12,13]. This line of research was pursued [7] with Haar functions on a very simple genetic algorithm using binary coding of integers. In the limit of infinite sampling, the sampling probability can be related to wavelet analysis and consequently to filter theory. Markov-based multiresolution search methods are new techniques [14] that were designed as a generalization of the above-mentioned algorithm to both discrete and continuous search spaces. They are suited to problems for which target samples are found in regions of high average fitness values. In an adaptive multiresolution search, high average fitness regions are sampled, on average, more often than low average fitness ones by making the sampling range dependent on the fitness of the last sample. If a high fitness element is found, the next sample is chosen with a high probability within a short range of the previous one. On the contrary, if a low fitness element is obtained, the next sample is chosen preferentially within a large range. In strong contrast to other multiresolution adaptive search techniques, such as multiresolution simulated annealing [15], multiresolution genetic algorithms [8] or multiresolution Monte Carlo Markov chains [16], the dyadic structure of the algorithms discussed in this article permits the easy computation of the discovery probability of a target element. Synergies between Markov theory and multiresolution analysis can be exploited to estimate for quite large problems the outcome of the search. The Markov transition matrices associated to a multiresolution search have a sparse structure that reduces considerably the necessary computing power to estimate the discovery probabilities of a solution based on the knowledge of the fitness function. Fig. 1 shows the general form of the algorithm. At each iteration, a resolution m is associated to the candidate solution x_{input}. The resolution m is chosen with a probability which is a function of the fitness $f(x)$:

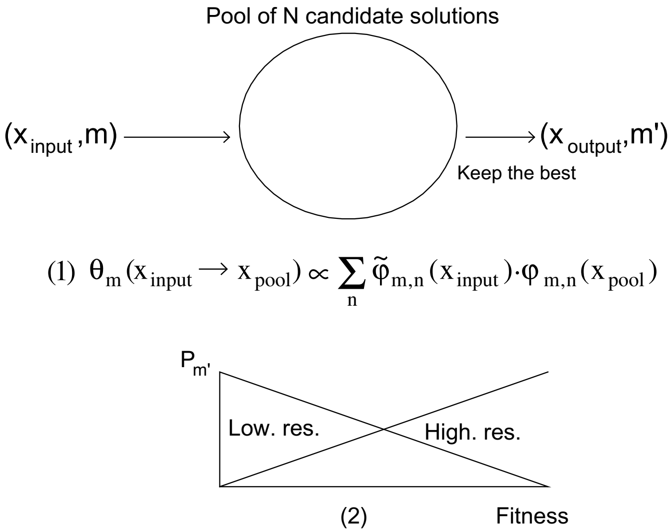


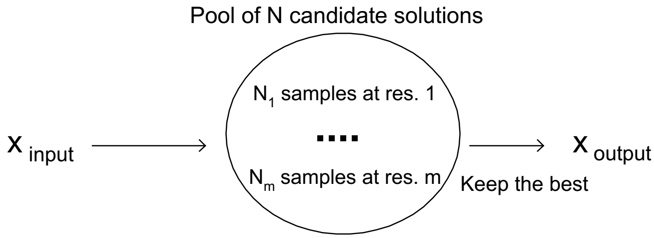
Fig. 1. In a Markov-based multiresolution search, N candidate solutions are chosen according to the probability distribution (1) and the best candidate is kept. The resolution m' is chosen with probability Pm' given by (2). In this example, only two resolutions (low and high) are taken.

$\mathrm{Prob}(m) = P_m(f(x))$. An intermediary pool of N candidate solutions is sampled using the probability distribution $\theta(x_{\mathrm{input}} \to x_{\mathrm{pool}})$. The probability distribution $\theta(x_{\mathrm{input}} \to x_{\mathrm{pool}})$ describes the probability of sampling the element x_{pool} when x_{input} is the last winning sample. The sampling probability distribution is chosen proportional to $\sum_n \tilde{\varphi}_{m,n}(x_{\mathrm{input}}) \cdot \varphi_{m,n}(x_{\mathrm{pool}})$

$$\theta(x_{\mathrm{input}} \to x_{\mathrm{pool}}) \propto \sum_n \tilde{\varphi}_{m,n}(x_{\mathrm{input}}) \cdot \varphi_{m,n}(x_{\mathrm{pool}}) \tag{1}$$

with $\varphi_{m,n}(x) = \varphi((x-n)/2^m)$, $\tilde{\varphi}_{m,n}(x) = \tilde{\varphi}((x-n)/2^m)$ and m, n integer. Finally, the candidate solution x_{output} with the largest fitness value is kept as input for the next generation. In order to prevent the ejection of the search from a promising region after sampling a single low fitness element, a number N of elements are sampled at each iteration step.

In principle, the only requirement on the scaling functions is that the resulting probability distribution is always positive. In wavelet-based search methods, $\tilde{\varphi}$ is the dual scaling function. In that case, Eq. (1) is the low-resolution kernel of a wavelet decomposition [17]. For orthogonal wavelets $\tilde{\varphi}_{m,n} = \varphi_{m,n}$. Except for Haar scaling functions, all scaling functions have negative values. In order to guarantee a positive probability distribution, several approaches may be used. For instance, a minimal value of the low-resolution sampling probability, corresponding to a random search with a uniform distribution, may guarantee a positive probability distribution.



$$\theta_m(x_{input} \to x_{pool}) \propto \sum_n \tilde{\varphi}_{m,n}(x_{input}) \cdot \varphi_{m,n}(x_{pool})$$

Fig. 2. In a Markov-based multiresolution search for which the fitness is either not bounded to one or only qualitatively known, the multiresolution search algorithm consists of choosing at each resolution a fixed number N_m of candidates with the probability distribution θ_m. At each iteration, the best candidate is kept.

The general multiresolution search explained in Fig. 1 assumes that the quality of the sample can be defined by a value between zero and one. Such a fitness function is not always available. If the quality of the sample cannot be quantified by a value between zero and one, then the algorithm must be slightly adapted. The algorithm presented in Fig. 2 is such an adaptation. The algorithm is a special case of the multiresolution search approach (Fig. 1) with a sampling probability independent of the fitness value.

The algorithm in Fig. 2 can be used in situations for which only a qualitative characterization of the sample is possible ("good", "very good", ...).

2.1. Limits of the Markov-based multiresolution search methods

In any adaptive search method, an important information is the maximal improvement of the search method compared to a random search. In a multiresolution search method, a lower bound to the maximal acceleration of the search compared to a random search can be given. In order to find that bound, fitness functions that are particularly adapted to a multiresolution search are constructed. Fig. 3 shows an example for a 2-resolution search using Haar scaling functions. The expected number of samples to reach the first level on the fitness function (see Fig. 3) is $\sqrt{\text{Size}}$ with Size the number of elements in the search space. Once on the first level, the expected number of samples to reach the target element is again $\sqrt{\text{Size}}$. Neglecting the probability of sampling directly the target sample without first passing by the first level, the expected number of samples is $2 \cdot \sqrt{\text{Size}}$. The algorithm described in Fig. 2 exploits best the available prior information and uses the best strategy ("a random search") within each level. Repeating the same reasoning with m resolutions, the minimum expected number of samples is

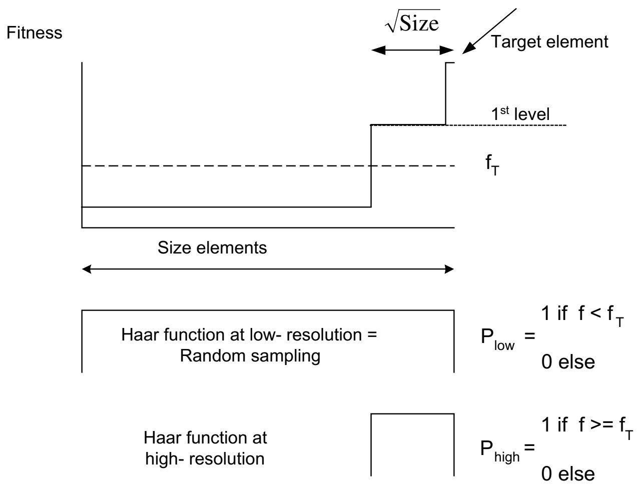


Fig. 3. The above fitness function is almost optimal for a 2-resolution search in the sense that the expected number of samples to discover the sample is close to the minimum value obtainable with a 2-resolution search on Size elements (for a large search space!). The expected number of sample to discover the target solution is about $2 \cdot \sqrt{\text{Size}}$. For a m-resolution search the expected number of samples is about $m\sqrt[m]{\text{Size}}$. For a search space with 10^6 elements, it represents, compared to a random search, an improvement by a factor 500 for a 2-resolution search and about 3300 for a 3-resolution search.

proportional to $m \cdot \sqrt[m]{\text{Size}}$. In summary, the potential of the multiresolution search method is a function of the number of resolutions: the higher the number of resolutions, the higher the maximum gain is in comparison to a random search. (Let us notice that considering the ensemble of all search problems in a search space with Size elements, the more levels one uses, the smaller is the probability that a problem fulfills the prior information!) Interestingly, if the quality of the sample is only qualitatively known as in the situation of Fig. 2, then the minimum expected number of samples is not much higher: $m^2 \cdot \sqrt[m]{\text{Size}}$.

Multiresolution search can be easily generalized to p variables by using in Eq. (1) multivariable mother functions: $\varphi_{m,n}(x_1,\dots,x_p)$ and $\tilde{\varphi}_{m,n}(x_1,\dots,x_p)$. Without further assumptions, the expected number of samples to discover a target sample is at least of the order of $m \cdot \text{Size}^{1/m}$ (m: number of resolution levels, Size: dimension of the search) [14]. This number can be reduced if some relationship does exist between the variables. If the fitness function is for instance separable, $f(x_1,\dots,x_p) = \sum_{i=1}^{p} f_i(x_i)$, the search problem can be transformed into p independent search problems. In that case, the expected number of samples to discover a target sample within a search space may be as low as about $m \cdot p \cdot \text{Size}^{1/(m \cdot p)}$.

2.2. Choosing the right resolution

The main measure of the quality of a stochastic search algorithm on some problem is given by the probability of discovering an acceptable solution at iteration k. For a Markov process, the discovery probability can be computed using the associated transition matrix [18]. The often-sparse structure of the transition matrix allows the computation of the discovery probability as a function of the number of iterations for quite large problems. The possibility of computing the performance of a Markov-based multiresolution search on different fitness functions is extremely valuable. So one can gain some experience on how to choose the search's parameters for the search. In a 2-resolution Markov-based search using a random sampling at low resolution, the only free parameter after the basis function has been chosen is the choice of the high-resolution level. Fig. 4 shows this with a very simple, but quite representative, example, a 1-sample search based on Haar functions. The results are quite typical and bear therefore some generality.

At a too low-resolution, the search is less efficient than the optimal search but is better than a random search. At a too high resolution, the search converges (too) rapidly towards some high-average fitness region. After a number of iterations, the search becomes less efficient than a random search, as the search becomes trapped in a high but not optimal region. In other words, at a too high resolution, it is better to restart the search than to persist too long in an unsuccessful search. The best resolution for the search in Fig. 4 corresponds to the characteristic size of the fitness function.

3. Wavelet-based search using an estimator approach

In this section, new adaptive wavelet-based search methods are introduced. We will show that by using the properties of the mother scaling functions associated to a wavelet decomposition, the sampling probability distribution can be made proportional to a low-resolution version of the fitness function. In this new approach to search, the learning and the exploitation phase are not separated as in estimation of density approaches [19], a significant advantage. From the memory point of view, the method requires only the storage of two values per low-pass coefficient.

The basic assumption beyond multiresolution and wavelet-based search is that at some scale, target elements are found with a high probability in regions of high average fitness. A way to exploit this information is to sample the search space proportionally to the smoothed fitness function. Concretely the sampling probability distribution S is chosen proportional to $\hat{f}$ with

$$\hat{f}(x) = \sum_m P_m(f) \cdot \sum_n \hat{c}_{m,n}(f) \cdot \varphi_{m,n}(x) \tag{2}$$

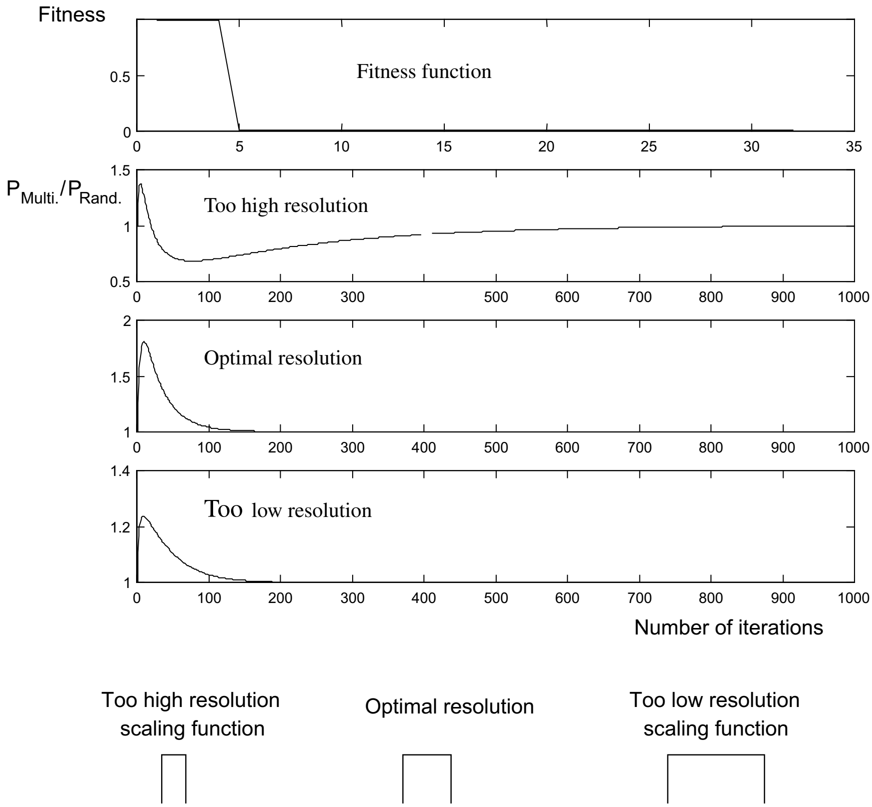


Fig. 4. Depending on the choice of the resolution, the search quality may change considerably. The curves show 3 examples of the normalized cumulative probability of discovering the highest fitness element ($x_s = 1$) (P_{Multi}: cumulative probability of discovering the highest fitness element; P_{Rand}: cumulative probability with a random search). A value above one means that the cumulative probability of discovering the target sample is higher than with a random search. For a large number of iterations, the cumulative probability tends to one and therefore the ratio tends asymptotically to one. The Haar scaling functions used at the highest-level of resolution are shown below the curves (two-resolution levels search with $P_{\mathrm{low}} = (1 - f)$ and $P_{\mathrm{high}} = f$). The first level of resolution corresponds to a random search.

The coefficients $\hat{c}_{m,n}$ are estimated with a wavelet network or a wavelet estimator [7]. In the later case, the coefficients are actualised with the following equation

$$\hat{c}_{m,n}(f) = \sum_i f(x_i) \cdot \tilde{\varphi}(x_i) \Big/ \sum_i \tilde{\varphi}(x_i). \tag{3}$$

in which $\tilde{\varphi}(x)$ is the dual scaling function. Fig. 5 shows with an example that the wavelet-based search has, at equilibrium, a number of interesting features:

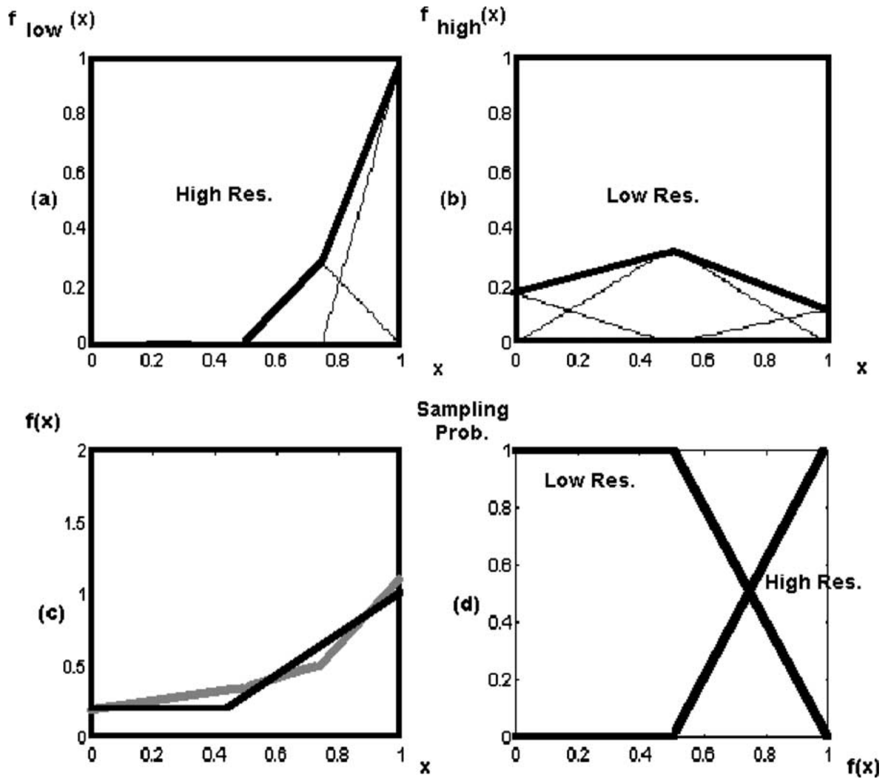


Fig. 5. The function (Fig. 5c, black curve) was sampled at two resolutions (Fig. 5d) using Eqs. (2) and (3) and biorthogonal 4.2 splines. The sampling probability at high-resolution is proportional, at equilibrium, to f_{high} (Fig. 5a), with f_{high} a weighted sum of triangular functions. The low-resolution sampling probability distribution is proportional to f_{low} (Fig. 5b). The resulting sampling probability distribution is proportional to an estimation (Fig. 5c, grey line) of the original function.

- The sampling probability distribution is proportional to an approximation of the fitness function.
- Regions in the search space with a high average fitness are sampled more often than low average fitness regions.
- The sampling probability distribution is computed at a higher resolution in high average fitness regions.

3.1. One-resolution wavelet-based search

The equations in the previous subsection are probably best explained if one starts by describing the 1-resolution case, corresponding to setting $P_m = 1$ for some resolution m. The sampling probability distribution S is then proportional to $\hat{f}(x)$:

$$S(x) \propto \hat{f}(x) = \sum_n \hat{c}_{m,n}(f) \cdot \varphi_{m,n}(x) \tag{4}$$

or equivalently $\hat{c}_{m,n}(f) \propto c_{m,n}(S)$ with $S(x) = \sum_n c_{m,n}(S) \cdot \varphi_{m,n}(x)$.

The search is practically implemented using Eqs. (3) and (4). The coefficients $\hat{c}_{m,n}(f)$ are computed at each iteration with Eq. (3) summing over all past samples x_i and the sampling probability distribution $S(x)$ is chosen proportional to $\sum_n \hat{c}_{m,n}(f) \cdot \varphi_{m,n}(x)$.

Contrarily to the Markov-based approach in Section 2, the discovery probability cannot be easily estimated. The sampling probability distribution can only be estimated at equilibrium. The coefficients $\hat{c}(f)$ at equilibrium are related to the function f [1]

$$(\hat{c}_{m,n}(f))^2 = (c_{m,n}(\hat{f}))^2 \cong c_{m,n}(f \cdot \hat{f}) \tag{5}$$

The low-resolution projection coefficient $c_{m,n}(f \cdot \hat{f})$ of the product $f(x) \cdot \hat{f}(x)$ is equal to $(\hat{c}_{m,n}(f))^2$ with $c_{m,n}(\hat{f})$ corresponding to the low-resolution estimation of the fitness function obtained with Eq. (3). For Haar wavelets, one shows that $\hat{c}_{m,n}(f) = c_{m,n}(f)$. The expected distribution of samples is therefore in that case proportional to the low-passed function $\hat{f}(x) = \sum c_{m,n}(x) \cdot H_{m,n}(x)$. For splines or other mother scaling functions, the probability distribution function is proportional to a low-resolution version $\hat{f}$ of the fitness function satisfying Eq. (5).

A conceptually important special case is when splines are chosen as scaling mother functions. In that case a fuzzy interpretation of the results can be given within the framework of the Takagi–Sugeno model. Spline-based adaptive search methods permit to extend to search and optimization problems the fuzzy-wavelet methodologies used in estimation theory [7]. Fig. 6 shows an example using biorthogonal 4.2 splines [17]. The search furnishes, beside a list of high-fitness data points, a fuzzy representation of the fitness function. One obtains expressions of the kind:

$$\textit{if X is Large than fitness is Large } (C) \tag{6}$$

with C the confidence level (using a center of gravity defuzzification, rules of the form of Eq. (6) can be described within the framework of the Takagi–Sugeno model [7]).

[1] Proof of Eq. (5): at equilibrium, one obtains from Eq. (4) that

$$\sum_{\substack{\text{regular}\\\text{sampling}}} f(x) \cdot \hat{f}(x) \cdot \tilde{\varphi}(x) \Big/ \sum \hat{f}(x) \cdot \tilde{\varphi}(x) \cong \sum_{\substack{\text{regular}\\\text{sampling}}} \hat{f}(x) \cdot \tilde{\varphi}(x) \Big/ \sum \tilde{\varphi}(x)$$

which after reorganization of the terms furnishes the last equality.

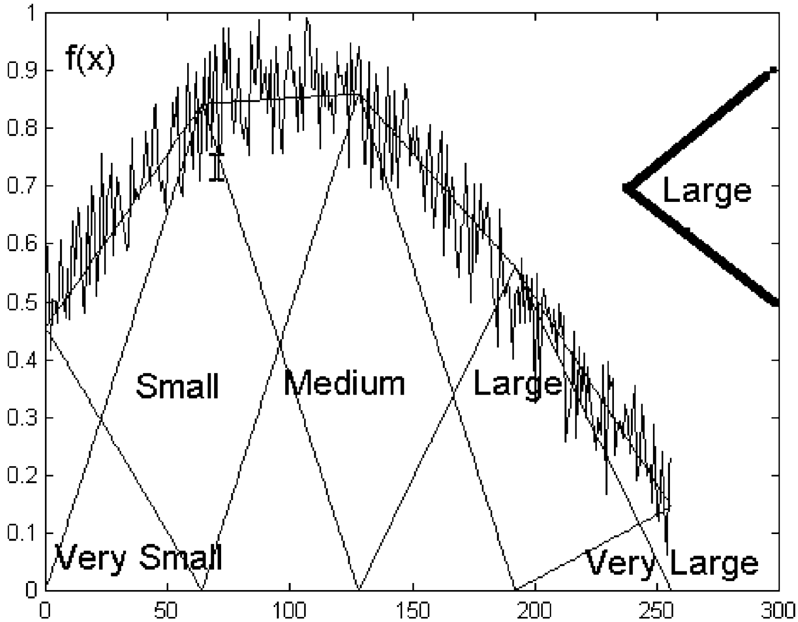


Fig. 6. Spline-based multiresolution search using biorthogonal 4.2 splines. The sampling probability $P_m(x)$ at equilibrium is proportional to $\hat{f}(x)$ (—). The search results can be put under a fuzzy form.

A significant advantage of the fuzzy approach is that information on the fitness can be easily introduced beforehand by initialising some coefficients in Eq. (3).

The wavelet-based search can be quite easily generalized to higher dimensional projections. At high dimensions, the problem is that the number of trials to discover a solution on a multivariable fitness function increases quite rapidly with the number of variables. The expected number of trials can be quite significantly reduced, if there exists some low-dimension projections on which the fitness function can be decomposed. In the next subsection, one examines how the information that the fitness function is separable can be exploited. In order to simplify the notation, we will deal with the special case of a fitness function that can be decomposed into the sum of functions of one variables

$$f(x_1, \ldots, x_j, \ldots, x_p) = \sum_{j=1,\ldots,p} g_j(x_j) \tag{7}$$

At each step, the sampling probability is chosen proportional to

$$S(x_1, \ldots, x_p) \propto \prod_{j=1,\ldots,p} \hat{g}_j(x_j) \tag{8}$$

with $\hat{g}_j(x) = \sum_n \hat{c}_{m,n}(g_j) \cdot \varphi_{m,n}(x_j)$.

The coefficient $\hat{c}_{m,n}$ are estimated using a generalization of the one-dimensional case:

$$\hat{c}_{m,n}(g_j) = \sum_k f(x_1(k), \ldots, x_p(k)) \cdot \tilde{\varphi}_{m,n}(x_j(k)) \Big/ \sum_k \tilde{\varphi}_{m,n}(x_j(k)) \tag{9}$$

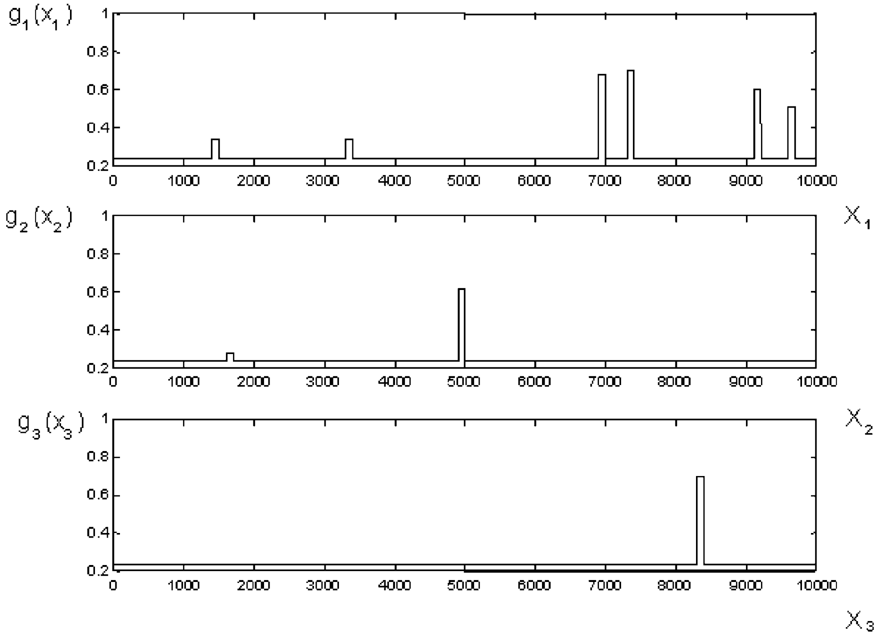
At equilibrium, the sampling probability can be shown to be proportional to

$$\hat{c}_{m,n}(g_j) \propto \sum_{a \neq j} \overline{g_a} + \langle g_j \cdot \hat{g}_j, \tilde{\varphi}_{m,n} \rangle \Big/ \langle \hat{g}_j, \tilde{\varphi}_{m,n} \rangle \tag{10}$$

with $\overline{g_a}$ the average of g_a over the whole search space. (The following definition is used: $\langle a, b \rangle = \int a(x) \cdot b(x) \cdot \mathrm{d}x$.).

Except for the constant $\overline{g_a}$ the estimated fitness function at equilibrium is of the same form as in the 1-variable case. For separable functions, the number of samples at equilibrium is therefore related to the fitness function. The constant sets a bound to the maximal gain in the search time. Numerical experiments with functions satisfying the basic assumption behind multiresolution search methods shows that Markov-based multiresolution search are often more efficient than wavelet-based search with memory to discover a target element. The wavelet-based search method has a number of properties that the Markov-based approach does not have. In addition to giving a solution to the search problem, it furnishes an approximation of the fitness function, an approximation that may be quite useful in adaptive systems for instance to estimate the stability of the proposed solution. A large difference between the estimation of the fitness function and the best solution is an indication that the solution may not be very stable. Also in the fuzzy version of the algorithm, qualitative information on the fitness function can be easily included prior to the search. This information is refined during the search process as more data are sampled.

The results are much better if after a number of iterations, the search is focused on the region of highest fitness. Fig. 7 shows a prototypic example in which high values of the fitness are located in a limited number of small regions with some well-defined characteristic dimension. The results obtained with a wavelet-based search with memory (the search is restricted to the highest fitness region after 2500 samples) are compared to a Markov-based multiresolution search and a random search. Both multiresolution search methods use Haar functions of support equal to the characteristic size of the high-value regions of the fitness function. The 2-resolution Markov-based search (Fig. 2) is adapted to a 3-variable space. Four samples are drawn at each iteration, one at high resolution and one per variable at low resolution. (The one variable is sampled at low resolution, while the two other variables are sampled at high resolution.) At each iteration, the best sample is the winner. Both the Markov-based and the multiresolution search with memory are by far better than a random search. On average, the Markov-based search is better than the search with memory.



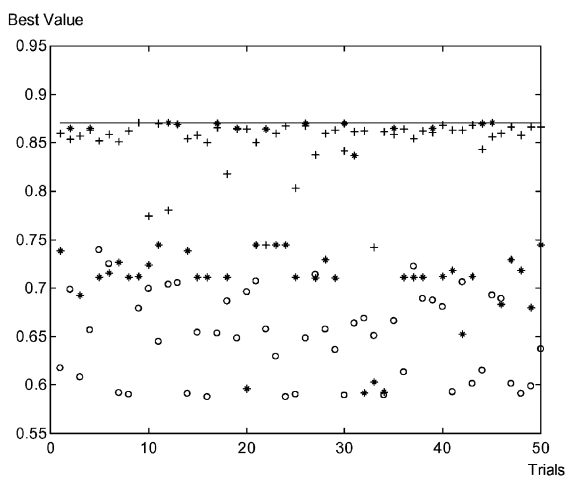


Fig. 7. The surface given by $f(x_1,x_2,x_3) = 1/3(g_1(x_1) + g_2(x_2) + g_3(x_3)) + W(x_1,x_2,x_3)$ with g_1, g_2, g_3 shown in the top curves and $W(x_1,x_2,x_3)$ a term from a random uniform distribution between [0, 0.2], was searched for high values with different methods: Haar-based search with memory in which the search is restricted to the highest fitness region after 2500 samples (cross), Markov-based 2-resolution search (star), random search (circle). The results are shown after 3000 samples. Each point corresponds to one trial. Both multiresolution search methods are on average much better than a random search. The black line gives the largest value.

4. Conclusions

Multiresolution search may considerably outperform random search provided the prior information justifies directing the search towards regions with high local average fitness. In other words, multiresolution search methods are suited to problems for which target samples are found in regions of high average fitness values. Two search algorithms, a Markov-based multiresolution search and a wavelet-based multiresolution search using elements of estimation theory, have been studied. Besides its simplicity to implement, the Markov-based multiresolution search has the great advantage to allow "reverse-engineering". Starting from a known fitness function the discovery probability of a target sample can be computed for quite large search spaces. The Markov-based approach is very simple and can be applied to very large search spaces. A 2-resolution search works best if the fitness function has low values on most of the search space and acceptable solutions are found in small clusters of high average values. Rescaling of the fitness function can achieve this, provided that rescaling preserves the prior information. If the fitness function are (almost) separable then a 2-resolution search on each variable will already lead to a quite large improvement compared to a random search as the gain on each variable (almost) factors. In the wavelet-based multiresolution search method using estimation theory, the sampling probability distribution is constructed so as to be a low-resolution estimation of the fitness function. At equilibrium, a simple relation exists between the sampling probability distribution and the fitness function providing a natural connection between estimation and search-and-optimization theory. This approach is particularly recommended if the estimation of the fitness function serves other purposes, for instance to estimate the stability of a the solution or as a potential source of new solutions in a non-stationary system. It requires the storage of only a small number of coefficients to store an estimation of the fitness function. The search is often less efficient than the Markov approach. In conjunction to another search method, the algorithm can however become quite efficient. The multiresolution search is used to localize the regions of interest, while the second algorithm focus on those interesting regions.

References

[1] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Trans. Evolut. Comput. 1 (1997) 67–82.

[2] J.H. Holland, Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.

[3] B. Selman, D.G. Levesque, A. Mitchell, A new method for solving hard satisfiability problems, in: Proceedings of AAAI-92, San Jose, CA, 1992, pp. 440–446.

[4] M. Dorigo, G. Di Caro, L.M. Gambardella, Ant algorithm for discrete optimization, Artif. Life 5 (1999) 137–172.
[5] R. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, 1957.
[6] Z. Michalewicz, D.B. Fogel, How to Solve It: Modern Heuristics, Springer, Berlin, 2000.
[7] M. Thuillard, in: Wavelets in Soft Computing, World Scientific Series in Robotics and Intelligent Systems, vol. 25, World Scientific, Singapore, 2000.
[8] L. Voicu, W.A. Rabadi, Object support reconstruction of its autocorrelation using multiresolution genetic algorithms, Opt. Eng. 36 (1997) 2280–2827.
[9] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[10] M.D. Vose, The Simple Genetic Algorithm: Foundations and Theory, MIT Press, Boston, 1999.
[11] A.D. Bethke, Genetic algorithms as function optimizers, Ph.D. Work, Michigan University, 1981.
[12] D.E. Goldberg, Genetic algorithms and Walsh functions: part I, a gentle introduction, Complex Syst. 3 (1989) 129–152.
[13] D.E. Goldberg, Genetic Algorithms, Addison-Wesley, USA, 1991.
[14] M. Thuillard, Adaptive multiresolution and wavelet-based search methods, in: Proceedings of First International IEEE Symposium on Intelligent Systems, Varna, Bulgaria, 10–12 September 2002, vol. I, 2002, pp. 110–114.
[15] S. Loncaric, Z. Majcenic, Multiresolution simulated annealing for brain image analysis, in: Proceedings of SPIE Medical Imaging, San Diego, USA, vol. 3661, 1999, pp. 1139–1146.
[16] C.H. Holloman, H.K.H. Lee, D.M. Higdon, Multi-resolution genetic algorithms and Markov chain Monte Carlo, ISDS, Duke, Discussion Paper Series 02–06 (2002).
[17] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.
[18] F.R. Gantmacher, The Theory of Matrices, Chelsea, New York, 1974.
[19] H. Mühlenbein, G. Pass, From recombination of genes to the estimation of distributions I. Binary parameters, in: Parallel Problem Solving from Nature, in: H.-M. Voigt, W. Ebeling, I. Rechenberg, H.-P. Schwefel (Eds.), Lecture Notes in Computer Science, 1141, Springer-Verlag, 1996, pp. 178–187.