Journal of Mathematical Modelling and Algorithms **3:** 313–322, 2004.

A No-Free-Lunch Theorem for Non-Uniform Distributions of Target Functions

CHRISTIAN IGEL and MARC TOUSSAINT
Chair of Theoretical Biology, Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany. e-mail: {christian.igel,marc.toussaint}@neuroinformatik.rub.de

Abstract. The sharpened No-Free-Lunch-theorem (NFL-theorem) states that, regardless of the performance measure, the performance of all optimization algorithms averaged uniformly over any finite set F of functions is equal if and only if F is closed under permutation (c.u.p.). In this paper, we first summarize some consequences of this theorem, which have been proven recently: The number of subsets c.u.p. can be neglected compared to the total number of possible subsets. In particular, problem classes relevant in practice are not likely to be c.u.p. The average number of evaluations needed to find a desirable (e.g., optimal) solution can be calculated independent of the optimization algorithm in certain scenarios. Second, as the main result, the NFL-theorem is extended. Necessary and sufficient conditions for NFL-results to hold are given for arbitrary distributions of target functions. This yields the most general NFL-theorem for optimization presented so far.

Mathematics Subject Classifications (2000): 90C27, 68T20.

Key words: evolutionary computation, No-Free-Lunch theorem.

1. Introduction

Search heuristics such as evolutionary algorithms, grid search, simulated annealing, and tabu search are general in the sense that they can be applied to any target function $f: \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ denotes the search space and $\mathcal{Y}$ is a set of totally ordered cost-values. Much research is spent on developing search heuristics that are superior to others when the target functions belong to a certain class of problems. But under which conditions can one search method be better than another? The No-Free-Lunch-theorem for optimization (NFL-theorem) roughly speaking states that all non-repeating search algorithms have the same mean performance when averaged uniformly over *all* possible objective functions $f: \mathcal{X} \to \mathcal{Y}$ [14, 9, 2, 15, 8, 1]. Of course, in practice an algorithm need not perform well on all possible functions, but only on a subset that arises from the real-world problems at hand. Recently, a sharpened version of the NFL-theorem has been proven that states that NFL-results hold (i.e., the mean performance of all search algorithms is equal) for any subset F of the set of all possible functions if and only if F is closed under permutation (c.u.p.) and each target function in F is equally likely [11].

In this paper, we address the following basic questions: When all algorithms have the same mean performance – how long does it take on average to find a desirable solution? How likely is it that a randomly chosen subset of functions is c.u.p., i.e., fulfills the prerequisites of the sharpened NFL-theorem? Do constraints relevant in practice lead to classes of target functions that are c.u.p.? And finally: How can the NFL-theorem be extended to non-uniform distributions of target functions? These questions can be accessed from the sharpened NFL-theorem. Answers are given in the following sections, after a formal description of the scenario considered in NFL-theorems.

2. Preliminaries

A finite search space $\mathcal{X}$ and a finite set of cost-values $\mathcal{Y}$ are presumed. Let $\mathcal{F}$ be the set of all objective functions $f: \mathcal{X} \to \mathcal{Y}$ to be optimized (also called target, fitness, energy, or cost functions). NFL-theorems are concerned with non-repeating black-box search algorithms (referred to as algorithms) that choose a new exploration point in the search space depending on the history of prior explorations: The sequence $T_m = \langle (x_1, f(x_1)), (x_2, f(x_2)), \ldots, (x_m, f(x_m)) \rangle$ represents m pairs of different search points $x_i \in \mathcal{X}$, $\forall i, j: x_i \neq x_j$ and their cost-values $f(x_i) \in \mathcal{Y}$. An algorithm a appends a pair $(x_{m+1}, f(x_{m+1}))$ to this sequence by mapping T_m to a new point x_{m+1}, $\forall i: x_{m+1} \neq x_i$. In many search heuristics, such as evolutionary algorithms or simulated annealing in their canonical form, it is not ensured that a point in the search space is evaluated only once. However, these algorithms can become non-repeating when they are coupled with a search-point database, see [4] for an example in the field of structure optimization of neural networks.

We assume that the performance of an algorithm a after m iterations with respect to a function f depends only on the sequence

$$Y(f, m, a) = \langle f(x_1), f(x_2), \ldots, f(x_m) \rangle$$

of cost-values the algorithm has produced. Let the function c denote a performance measure mapping sequences of cost-values to the real numbers. For example, in the case of function minimization a performance measure that returns the minimum cost-value in the sequence could be a reasonable choice. See Figure 1 for a schema of the scenario assumed in NFL-theorems.

Using these definitions, the original NFL-theorem for optimization reads:

THEOREM 1 (NFL-theorem [15]). *For any two algorithms a and b, any $k \in \mathbb{R}$, any $m \in \{1, \ldots, |\mathcal{X}|\}$, and any performance measure c*

$$\sum_{f \in \mathcal{F}} \delta(k, c(Y(f, m, a))) = \sum_{f \in \mathcal{F}} \delta(k, c(Y(f, m, b))). \tag{1}$$

Herein, δ denotes the Kronecker function ($\delta(i, j) = 1$ if $i = j$, $\delta(i, j) = 0$ otherwise). Proofs can be found in [14, 15, 8]. This theorem implies that for any two

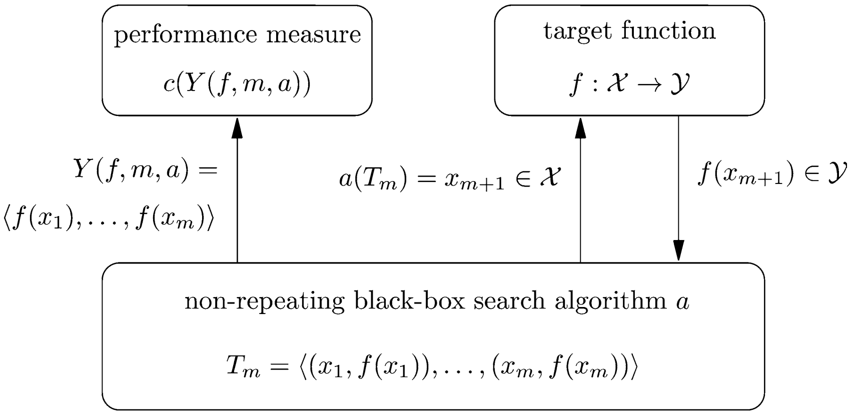

Figure 1. Scheme of the optimization scenario considered in NFL-theorems. A non-repeating black-box search algorithm a chooses a new exploration point in the search space depending on the sequence T_m of the already visited points with their corresponding cost-values. The target function f returns the cost-value of a candidate solution as the only information. The performance of a is determined using the performance measure c, which is a function of the sequence $Y(f, m, a)$ containing the cost-values of the visited points.

(deterministic or stochastic, cf. [1]) algorithms a and b and any function $f_a \in \mathcal{F}$, there is a function $f_b \in F$ on which b has the same performance as a on f_a. Hence, statements like "Averaged over all functions, my search algorithm is the best" are misconceptions. Note that the summation in (1) corresponds to uniformly averaging over all functions in $\mathcal{F}$, i.e., each function has the same probability to be the target function.

Recently, Theorem 1 has been extended to subsets of functions that are closed under permutation (c.u.p.). Let $\pi\colon \mathcal{X} \to \mathcal{X}$ be a permutation of $\mathcal{X}$. The set of all permutations of $\mathcal{X}$ is denoted by $\Pi(\mathcal{X})$. A set $F \subseteq \mathcal{F}$ is said to be c.u.p. if for any $\pi \in \Pi(\mathcal{X})$ and any function $f \in F$ the function $f \circ \pi$ is also in F.

EXAMPLE 1. Consider the mappings $\{0, 1\}^2 \to \{0, 1\}$, denoted by $f_0, f_1, \ldots, f_{15}$ as shown in Table I. Then the set $\{f_1, f_2, f_4, f_8\}$ is c.u.p., also $\{f_0, f_1, f_2, f_4, f_8\}$. The set $\{f_1, f_2, f_3, f_4, f_8\}$ is not c.u.p., because some functions are "missing", e.g., f_5, which results from f_3 by switching the elements $(0, 1)$ and $(1, 0)$.

In [10, 11] it is proven:

THEOREM 2 (sharpened NFL-theorem [10, 11]). *If and only if F is c.u.p., then for any two algorithms a and b, any $k \in \mathbb{R}$, any $m \in \{1, \ldots, |\mathcal{X}|\}$, and any performance measure c*

$$\sum_{f\in F} \delta(k, c(Y(f, m, a))) = \sum_{f\in F} \delta(k, c(Y(f, m, b))). \tag{2}$$

This is an important extension of Theorem 1, because it gives necessary and sufficient conditions for NFL-results for subsets of functions. But still Theorem 2

Table I. Functions $\{0, 1\}^2 \to \{0, 1\}$

(x_1, x_2)	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}
(0, 0)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
(0, 1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
(1, 0)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
(1, 1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

can only be applied if all elements in F have the same probability to be the target function, because the summations average uniformly over F. A variant of Theorem 1 giving sufficient, but not necessary conditions for NFL-results for more general distributions of the functions in $F \subseteq \mathcal{F}$ was already stated in [2], see Section 4 where we prove necessary and sufficient conditions for NFL-results to hold for arbitrary distributions of target functions.

However, note that violation of the precondition of the NFL-theorem does not lead to a “Free Lunch”, but nevertheless ensures the possibility of a “Free Appetizer”.

3. Recently Proven Implications of the Sharpened NFL

Theorem 2 tells us that on average all algorithms need the same time to find a desirable, say optimal, solution – but how long does it take? The average number of evaluations, i.e., the mean first hitting time, needed to find an optimum depends on the cardinality of the search space $|\mathcal{X}|$ and the number n of search points that are mapped to a desirable solution. As the set of *all* functions where n search points represent desirable solutions is c.u.p., it is sufficient to compute the average time to find one of these points for an arbitrary algorithm. It is given by

$$\frac{|\mathcal{X}| + 1}{n + 1}. \tag{3}$$

A proof can be found in [5], where this result is used to study the influence of neutrality (i.e., of non-injective genotype-phenotype mappings) on the time to find a desirable solution.

The NFL-theorems can be regarded as the basic skeleton of combinatorial optimization and are important for deriving theoretical results (e.g., see [5]). However, are the preconditions of the NFL-theorems ever fulfilled in practice? How likely is it that a randomly chosen subset is c.u.p.?

There exist $2^{(|\mathcal{Y}|^{|\mathcal{X}|})} - 1$ non-empty subsets of $\mathcal{F}$ and it holds:

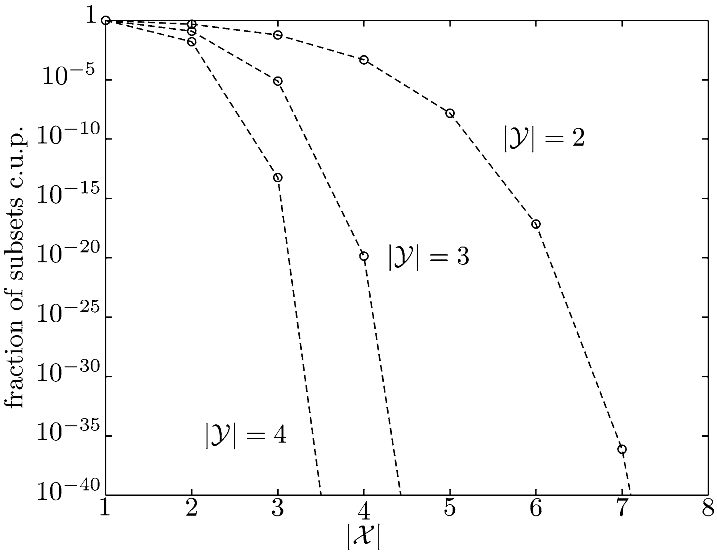

Figure 2. The ordinate gives the fraction of subsets closed under permutation on logarithmic scale given the cardinality of the search space $\mathcal{X}$. The different curves correspond to different cardinalities of the codomain $\mathcal{Y}$.

THEOREM 3 ([6]). *The number of non-empty subsets of* $\mathcal{Y}^{\mathcal{X}}$ *that are c.u.p. is given by*

$$2^{\binom{|\mathcal{X}|+|\mathcal{Y}|-1}{|\mathcal{X}|}} - 1 \tag{4}$$

and therefore the fraction of non-empty subsets c.u.p. is given by

$$\left(2^{\binom{|\mathcal{X}|+|\mathcal{Y}|-1}{|\mathcal{X}|}} - 1\right) \Big/ \left(2^{(|\mathcal{Y}|^{|\mathcal{X}|})} - 1\right). \tag{5}$$

A proof (using Lemma 1 quoted in Section 4) is given in [6].

Figure 2 shows a plot of the fraction of non-empty subsets c.u.p. versus the cardinality of $\mathcal{X}$ for different values of $|\mathcal{Y}|$. The fraction decreases for increasing $|\mathcal{X}|$ as well as for increasing $|\mathcal{Y}|$. More precisely, using bounds for binomial coefficients one can show that (5) converges to zero double exponentially fast with increasing $|\mathcal{X}|$ (for $|\mathcal{Y}| > e|\mathcal{X}|/(|\mathcal{X}| - e)$, where e is Euler's number). Already for small $|\mathcal{X}|$ and $|\mathcal{Y}|$ the fraction almost vanishes.

Thus, the statement "I'm only interested in a subset F of all possible functions and the precondition of the sharpened NFL-theorem is not fulfilled" is true with a probability close to one (if F is chosen uniformly and $\mathcal{Y}$ and $\mathcal{X}$ have reasonable cardinalities).

Although the fraction of subsets c.u.p. is close to zero already for small search and cost-value spaces, the absolute number of subsets c.u.p. grows rapidly with increasing $|\mathcal{X}|$ and $|\mathcal{Y}|$. What if these classes of functions are the relevant ones, i.e., those we are dealing with in practice?

Two assumptions can be made for most of the functions relevant in real-world optimization: First, the search space has some structure. Second, the set of objective functions fulfills some constraints defined based on this structure. More formally, there exists a non-trivial neighborhood relation on $\mathcal{X}$ based on which constraints on the set of functions under consideration are formulated, e.g., concepts like ruggedness or local optimality and constraints like upper bounds on the ruggedness or on the maximum number of local minima can be defined.

A neighborhood relation on $\mathcal{X}$ is a symmetric function $n: \mathcal{X} \times \mathcal{X} \to \{0, 1\}$. Two elements $x_i, x_j \in \mathcal{X}$ are called neighbors iff $n(x_i, x_j) = 1$. A neighborhood relation is called non-trivial iff $\exists x_i, x_j \in \mathcal{X}: x_i \neq x_j \wedge n(x_i, x_j) = 1$ and $\exists x_k, x_l \in \mathcal{X}: x_k \neq x_l \wedge n(x_k, x_l) = 0$. There are only two trivial neighborhood relations, either every two points are neighbored or no points are neighbored.

It holds:

THEOREM 4 ([6]). *A non-trivial neighborhood relation on $\mathcal{X}$ is not invariant under permutations of $\mathcal{X}$, i.e.,*

$$\exists x_i, x_j \in \mathcal{X}, \pi \in \Pi(\mathcal{X}): \quad n(x_i, x_j) \neq n(\pi(x_i), \pi(x_j)). \tag{6}$$

This result is quite general. Assume that the search space $\mathcal{X}$ can be decomposed as $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_l$, $l > 1$, and let on one component $\mathcal{X}_i$ exist a non-trivial neighborhood $n_i: \mathcal{X}_i \times \mathcal{X}_i \to \{0, 1\}$. This neighborhood induces a non-trivial neighborhood on $\mathcal{X}$, where two points are neighbored iff their i-th components are neighbored with respect to n_i – regardless of the other components. Thus, the constraints discussed below need only refer to a single component. Note that the neighborhood relation need not be the canonical one (e.g., the Hamming-distance for Boolean search spaces). For example, if integers are encoded by bit-strings, then the bit-strings can be defined as neighbored iff the corresponding integers are.

The following examples illustrate implications of Theorem 4.

EXAMPLE 2. Consider a non-empty subset $F \subset \mathcal{F}$ where the codomains of the functions have more than one element and a non-trivial neighborhood relation exists. If for each $f \in F$ it is not allowed that a global maximum is neighbored to a global minimum (i.e., we have a constraint “steepness”), then F is not c.u.p. – because for every $f \in F$ there exists a permutation that maps a global minimum and a global maximum of f to neighboring points.

EXAMPLE 3. Consider the number of local minima, which is often regarded as a measure of complexity [13]. Given a neighborhood relation on $\mathcal{X}$, a local minimum can be defined as a point whose neighbors all have worse fitness. Let $l(f)$ be the number of local minima of a function $f \in \mathcal{F}$ and let $B_{h_f} \subseteq \mathcal{F}$ be the set of functions in $\mathcal{F}$ with the same $\mathcal{Y}$-histogram as f (i.e., functions where the number of points in $\mathcal{X}$ that are mapped to each $\mathcal{Y}$-value is the same as for f, see Section 4 for details). Given a function f we define $l^{\max}(f) = \max_{f' \in B_{h_f}} l(f')$

as the maximal number of local minima that functions in $\mathcal{F}$ with the same $\mathcal{Y}$-histogram as f can possibly have. For a non-empty subset $F \subset \mathcal{F}$ we define $l^{\max}(F) = \max_{f \in F} l^{\max}(f)$. Let $g(F) \in \mathcal{F}$ be a function with $l(g(F)) = l^{\max}(F)$ local minima and the same $\mathcal{Y}$-histogram as a function in F. Now, if the number of local minima of every function $f \in F$ is constrained to be smaller than $l^{\max}(F)$ (i.e., $\max_{f \in F} l(f) < l^{\max}(F)$), then F is not c.u.p. – because $\exists f \in F$ with the same $\mathcal{Y}$-histogram as g and thus $\exists \pi \in \Pi(\mathcal{X})$: $f \circ \pi = g$. As a concrete example, consider all mappings $\{0, 1\}^\ell \to \{0, 1\}$ that have less than the maximum number of 2^{n-1} local minima w.r.t. the ordinary hypercube topology on $\{0, 1\}^\ell$. For example, this set does not contain mappings such as the parity function, which is one iff the number of ones in the input bitstring is even. This set is not c.u.p.

4. A Non-Uniform NFL-Theorem

In the sharpened NFL-theorem it is implicitly presumed that all functions in the subset F are equally likely since averaging is done by uniform summation over F. Here, we investigate the general case when every function $f \in \mathcal{F}$ has an arbitrary probability $p(f)$ to be the objective function. Such a non-uniform distribution of the functions in F appears to be much more realistic. Until now, there exist only very weak results for this general scenario. For example, let for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$p_x(y) := \sum_{f \in \mathcal{F}} p(f)\, \delta(f(x), y), \tag{7}$$

i.e., $p_x(y)$ denotes the probability that the search point x is mapped to the cost-value y. In [2, 3] it has been shown that a NFL-result holds if within a class of functions the function values are i.i.d. (independently and identically distributed), i.e., if

$$\forall x_1, x_2 \in \mathcal{X}: \quad p_{x_1} = p_{x_2} \quad \text{and} \quad p_{x_1,x_2} = p_{x_1} p_{x_2}, \tag{8}$$

where p_{x_1,x_2} is the joint probability distribution of the function values of the search points x_1 and x_2. However, this is not a necessary condition and applies only to extremely “unstructured” problem classes.

To derive a more general result let us introduce the concept of $\mathcal{Y}$-histograms. A *$\mathcal{Y}$-histogram* (*histogram* for short) is a mapping $h: \mathcal{Y} \to \mathbb{N}_0$ such that $\sum_{y \in \mathcal{Y}} h(y) = |\mathcal{X}|$. The set of all histograms is denoted $\mathcal{H}$. Any function $f: \mathcal{X} \to \mathcal{Y}$ implies the histogram $h_f(y) = |f^{-1}(y)|$ that counts the number of elements in $\mathcal{X}$ that are mapped to the same value $y \in \mathcal{Y}$ by f. Herein, $f^{-1}(y)$ returns the preimage $\{x \mid f(x) = y\}$ of y under f. Further, two functions f, g are called *h-equivalent* iff they have the same histogram. The corresponding h-equivalence class $B_h \subseteq \mathcal{F}$ containing all functions with histogram h is termed a *basis class*.

EXAMPLE 4. Consider the functions in Table I. The $\mathcal{Y}$-histogram of f_1 contains the value zero three times and the value one one time, i.e., we have $h_{f_1}(0) = 3$ and $h_{f_1}(1) = 1$. The mappings f_1, f_2, f_4, f_8 have the same $\mathcal{Y}$-histogram and are therefore in the same basis class $B_{h_{f_1}} = \{f_1, f_2, f_4, f_8\}$. It turns out that all basis classes are c.u.p. Furthermore, every set c.u.p. can be expressed as a union of basis classes. For example, the set $\{f_1, f_2, f_4, f_8, f_{15}\}$ is c.u.p. and corresponds to $B_{h_{f_1}} \cup B_{h_{f_{15}}}$.

It holds:

LEMMA 1 ([6]). (1) *Any subset $F \subseteq \mathcal{F}$ that is c.u.p. is uniquely defined by a union of pairwise disjoint basis classes.*

(2) *B_h is equal to the permutation orbit of any function f with histogram h, i.e., for all f*

$$B_{h_f} = \bigcup_{\pi \in \Pi(\mathcal{X})} \{f \circ \pi\}. \tag{9}$$

A proof is given in [6].

The following ramification of the sharpened NFL-theorem (derived independently in [7] and [12]) gives a necessary and sufficient condition for a NFL-result in the general case of non-uniform distributions:

THEOREM 5 (non-uniform sharpened NFL). *If and only if for all histograms h*

$$f, g \in B_h \Rightarrow p(f) = p(g), \tag{10}$$

then for any two algorithms a and b, any value $k \in \mathbb{R}$, any $m \in \{1, \dots, |\mathcal{X}|\}$, and any performance measure c

$$\sum_{f \in \mathcal{F}} p(f)\,\delta(k, c(Y(f, m, a))) = \sum_{f \in \mathcal{F}} p(f)\,\delta(k, c(Y(f, m, b))). \tag{11}$$

Proof. First, we show that (10) implies that (11) holds for any a, b, k, m, and c. It holds by Lemma 1(1)

$$\sum_{f \in \mathcal{F}} p(f)\,\delta(k, c(Y(f, m, a))) = \sum_{h \in \mathcal{H}} \sum_{f \in B_h} p(f)\delta(k, c(Y(f, m, a))) \tag{12}$$

using $f, g \in B_h \Rightarrow p(f) = p(g) = p_h$

$$= \sum_{h \in \mathcal{H}} p_h \sum_{f \in B_h} \delta(k, c(Y(f, m, a))) \tag{13}$$

as each B_h is c.u.p. we may use Theorem 2 proven in [10, 11]

$$= \sum_{h \in \mathcal{H}} p_h \sum_{f \in B_h} \delta(k, c(Y(f, m, b))) \tag{14}$$

$$= \sum_{f \in \mathcal{F}} p(f)\,\delta(k, c(Y(f, m, b))). \tag{15}$$

Now we prove that (11) being true for any a, b, k, m, and c implies (10) by showing that if (10) is not fulfilled then there exist a, b, k, m, and c such that (11) is also not valid. Let $f, g \in B_h$, $f \neq g$, $p(f) \neq p(g)$, and $g = f \circ \pi$. Let $\mathcal{X} = \{\xi_1, \ldots, \xi_n\}$. Let a be an algorithm that always enumerates the search space in the order $\xi_1, \ldots, \xi_n$ regardless of the observed cost-values and let b be an algorithm that enumerates the search space always in the order $\pi^{-1}(\xi_1), \ldots, \pi^{-1}(\xi_n)$. It holds $g(\pi^{-1}(\xi_i)) = f(\xi_i)$ for $i = 1, \ldots, n$ and $Y(f, n, a) = Y(g, n, b)$. We consider the performance measure

$$c^{\dagger}(\langle y_1, \ldots, y_m \rangle) = \begin{cases} 1 & \text{if } m = n \\ & \text{and } \langle y_1, \ldots, y_m \rangle = \langle f(\xi_1), \ldots, f(\xi_n) \rangle, \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

for any $y_1, \ldots, y_m \in \mathcal{Y}$. Then, for $m = n$ and $k = 1$, we have

$$\sum_{f' \in \mathcal{F}} p(f')\, \delta(k, c^{\dagger}(Y(f', n, a))) = p(f), \tag{17}$$

as $f' = f$ is the only function $f' \in \mathcal{F}$ that yields

$$\langle f'(\xi_1), \ldots, f'(\xi_n) \rangle = \langle f(\xi_1), \ldots, f(\xi_n) \rangle, \tag{18}$$

and

$$\sum_{f' \in \mathcal{F}} p(f') \delta(k, c^{\dagger}(Y(f', n, b))) = p(g), \tag{19}$$

and therefore (11) does not hold. □

The sufficient condition given in [3] is a special case of Theorem 5, because (8) implies

$$g = f \circ \pi \Rightarrow p(f) = p(g) \tag{20}$$

for any $f, g \in \mathcal{F}$ and $\pi \in \Pi(\mathcal{X})$, which in turn implies $g, f \in B_h \Rightarrow p(f) = p(g)$ due to Lemma 1(2).

The probability that a randomly chosen distribution over the set of objective functions fulfills the preconditions of Theorem 5 has measure zero. This means that in this general and realistic scenario the probability that the conditions for a NFL-result hold vanishes.

5. Conclusion

Recent studies on NFL-theorems for optimization were summarized and extended. As the main result, we derived necessary and sufficient conditions for NFL-results for arbitrary distributions of target functions and thereby presented the "sharpest" NFL theorem so far. It turns out that in this generalized scenario, the necessary conditions for NFL-results can not be expected to be fulfilled.

Acknowledgements

This work was supported by the DFG, grant Solesys-II SCHO 336/5-2. We thank Stefan Wiegand for fruitful discussions.

References

1. Droste, S., Jansen, T. and Wegener, I.: Optimization with randomized search heuristics – The (A)NFL theorem, realistic scenarios, and difficult functions, *Theoret. Comput. Sci.* **287**(1) (2002), 131–144.
2. English, T. M.: Evaluation of evolutionary and genetic optimizers: No free lunch, in L. J. Fogel, P. J. Angeline and T. Bäck (eds), *Proceedings of the Fifth Annual Conference on Evolutionary Programming (EP V)*, 1996, pp. 163–169.
3. English, T. M.: Optimization is easy and learning is hard in the typical function, in A. Zalzala, C. Fonseca, J.-H. Kim and A. Smith (eds), *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, 2000, pp. 924–931.
4. Igel, C. and Stagge, P.: Graph isomorphisms effect structure optimization of neural networks, in *International Joint Conference on Neural Networks (IJCNN 2002)*, 2002, pp. 142–147.
5. Igel, C. and Toussaint, M.: Neutrality and self-adaptation, *Natural Computing* **2**(2) (2003), 117–132.
6. Igel, C. and Toussaint, M.: On classes of functions for which no free lunch results hold, *Inform. Process. Lett.* **86**(6) (2003), 317–321.
7. Igel, C. and Toussaint, M.: Recent results on no-free-lunch theorems for optimization, arXiv preprint cs.NE/0303032, http://arxiv.org/abs/cs.NE/0303032, 2003.
8. Köppen, M., Wolpert, D. H. and Macready, W. G.: Remarks on a recent paper on the “No Free Lunch” theorems, *IEEE Trans. Evolut. Comput.* **5**(3) (1995), 295–296.
9. Radcliffe, N. J. and Surry, P. D.: Fundamental limitations on search algorithms: Evolutionary computing in perspective, in J. van Leeuwen (ed.), *Computer Science Today: Recent Trends and Development*, Lecture Notes in Comput. Sci. 1000, 1995, pp. 275–291.
10. Schumacher, C.: Fundamental limitations of search, Ph.D. thesis, University of Tennessee, 2000.
11. Schumacher, C., Vose, M. D. and Whitley, L. D.: The no free lunch and description length, in L. Spector, E. Goodman, A. Wu, W. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. Garzon and E. Burke (eds), *Genetic and Evolutionary Computation Conference (GECCO 2001)*, 2001, pp. 565–570.
12. Streeter, M. J.: Two broad classes of functions for which a no free lunch result does not hold, in E. Cantú-Paz, J. A. Foster, K. Deb, D. Davis, R. Roy, U.-M. O’Reilly, H.-G. Beyer, R. Standish, G. Kendall, S. Wilson, M. Harman, J. Wegener, D. Dasgupta, M. A. Potter, A. C. Schultz, K. Dowsland, N. Jonoska and J. Miller (eds), *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2003)*, 2003, pp. 1418–1430.
13. Whitley, D.: A free lunch proof for gray versus binary encodings, in W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela and R. E. Smith (eds), *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999)*, Vol. 1, 1999, pp. 726–733.
14. Wolpert, D. H. and Macready, W. G.: No free lunch theorems for search, Technical Report SFI-TR-05-010, Santa Fe Institute, Santa Fe, NM, USA, 1995.
15. Wolpert, D. H. and Macready, W. G.: No free lunch theorems for optimization, *IEEE Trans. Evolut. Comput.* **1**(1) (1997), 67–82.