No Free Lunch

Summer Lecture Series 2002

Thomas Jansen

`tjansen.gmu.edu or Thomas.Jansen@cs.uni-dortmund.de`

Overview

- “The Original” NFL Theorem
- An improved NFL Theorem
- NFL Assumptions
  - Objective Functions
  - Performance Measures
  - Optimization Scenarios
- NFL Implications

General Randomized Search Heuristics

Search in some search space X for points rated by a function f.

Black Box Scenario:

- no direct access to f
- no a priori knowledge
- only way to learn about f is sampling

Assumption: The heuristic is complete.

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \rightarrow Y$

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \to Y$
- X discrete

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \to Y$
- X discrete
- Y discrete

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \rightarrow Y$
- X discrete
- Y discrete
- X finite

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \rightarrow Y$
- X discrete
- Y discrete
- X finite
- Y finite

Optimizing an Arbitrary Function

What is an arbitrary function?

- $f : X \to Y$
- X discrete
- Y discrete
- X finite
- Y finite
- no function “more likely” than any other

Performance Measures

- function evaluations matter
- success measured in function values found

Performance Measures

- function evaluations matter
- performance measured in function val. found
- re-evaluations easy to avoid

Performance Measures

- function evaluations matter
- performance measured in function val. found
- re-evaluations easy to avoid

Consider algorithm A on function f and ignore re-evaluations.

Performance Measures (continued)

Consider algorithm's trace
$T(A, f, t) =< (x_1, f(x_1)), (x_2, (f(x_2)), \ldots, (x_t, f(x_t)) >$.

Performance Measures (continued)

Consider algorithm's trace
$T(A, f, t) =< (x_1, f(x_1)), (x_2, (f(x_2)), \ldots, (x_t, f(x_t)) >$.

Measure performance based on performance vector
$V(T(A, f, t)) =< f(x_1), f(x_2), \ldots, f(x_t) >$.

Performance Measures (continued)

Consider algorithm's trace
$T(A, f, t) =< (x_1, f(x_1)), (x_2, (f(x_2)), \ldots, (x_t, f(x_t)) >$.

Measure performance based on performance vector
$V(T(A, f, t)) =< f(x_1), f(x_2), \ldots, f(x_t) >$.

Performance measure
$M : \{V(T(A, f, t)) \mid A, f, t\} \rightarrow \mathbb{R}$

Performance Measures (continued)

Consider algorithm's trace
$T(A, f, t) =< (x_1, f(x_1)), (x_2, (f(x_2)), \ldots, (x_t, f(x_t)) >$.

Measure performance based on performance vector
$V(T(A, f, t)) =< f(x_1), f(x_2), \ldots, f(x_t) >$.

Performance measure
$M : \{V(T(A, f, t)) \mid A, f, t\} \rightarrow \mathbb{R}$

Global performance measure

$$\left(\sum_{f \in Y^X} M\left(V(T(A, f, |X|))\right) \right) / |Y|^{|X|}$$

The No Free Lunch Theorem

No Free Lunch Theorem:

For all finite sets X, Y

and for all global performance measures,

all search algorithms perform equal.

Proving the NFL Theorem (I)

First step: Consider only deterministic search algorithms.

Proving the NFL Theorem (I)

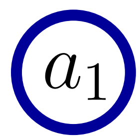
First step: Consider only deterministic search algorithms.

How can we describe a deterministic search algorithm?

Proving the NFL Theorem (I)

First step: Consider only deterministic search algorithms.

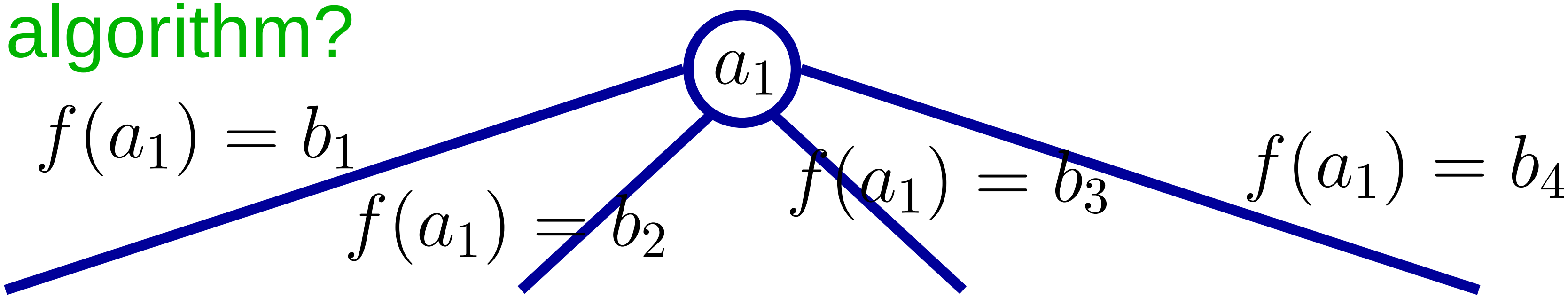
How can we describe a deterministic search algorithm?



Proving the NFL Theorem (I)

First step: Consider only deterministic search algorithms.

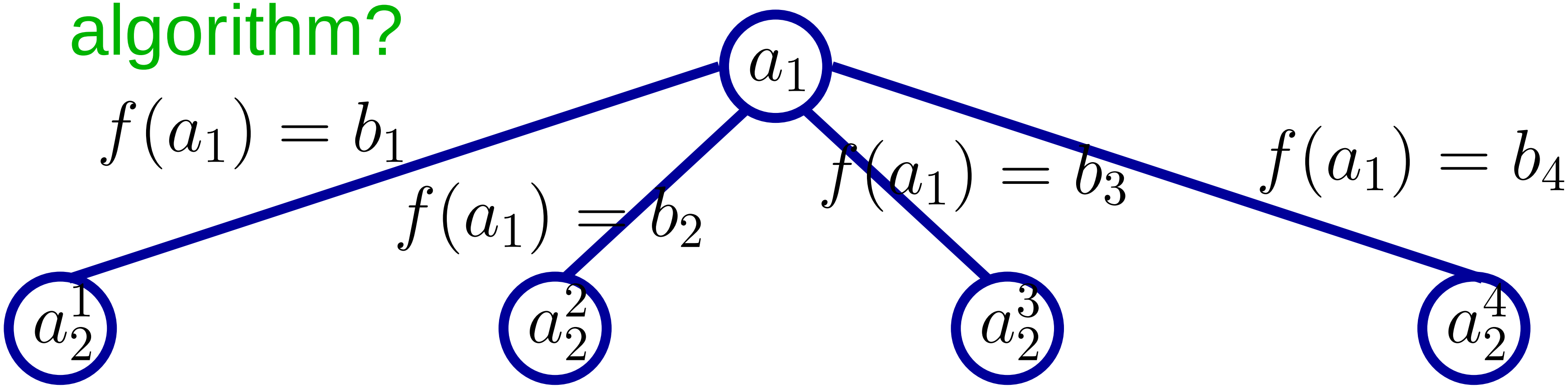
How can we describe a deterministic search algorithm?



Proving the NFL Theorem (I)

First step: Consider only deterministic search algorithms.

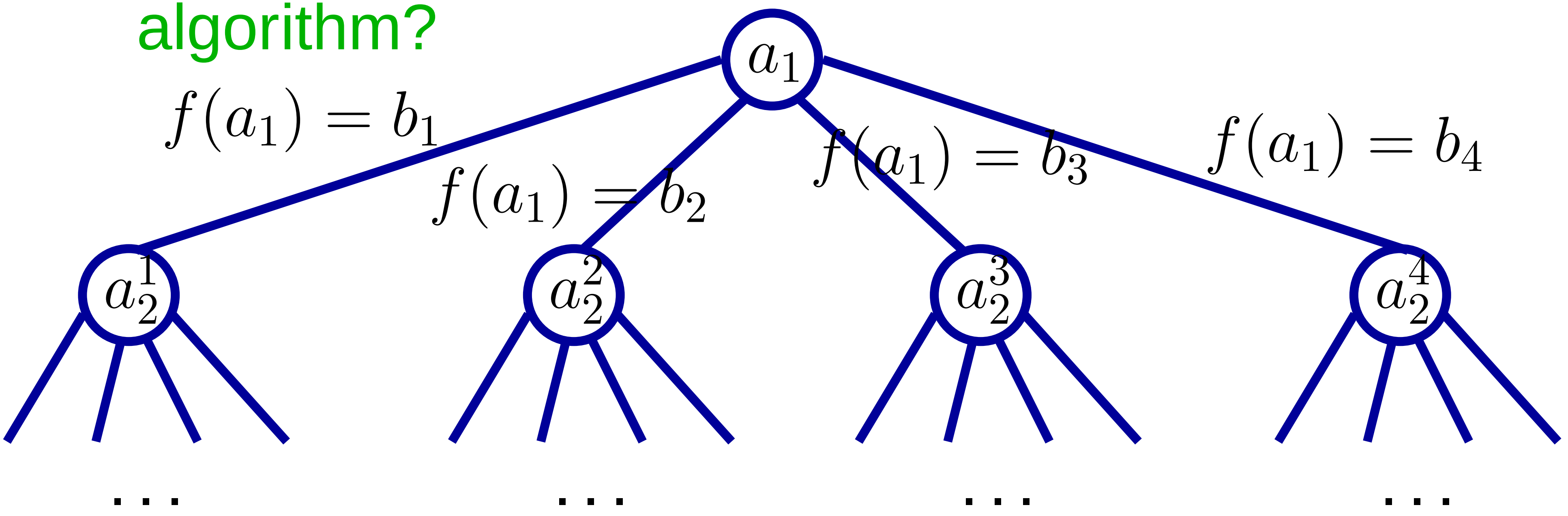
How can we describe a deterministic search algorithm?



Proving the NFL Theorem (I)

First step: Consider only deterministic search algorithms.

How can we describe a deterministic search algorithm?



Proving the NFL Theorem (II)

Consider a fixed search algorithm A.

Claim:

$$\forall f, g: \; V(T(A, f, |X|)) = V(T(A, g, |X|)) \Rightarrow f = g$$

Proving the NFL Theorem (II)

Consider a fixed search algorithm A.

Claim:

$$\forall f, g : \; V(T(A, f, |X|)) = V(T(A, g, |X|)) \Rightarrow f = g$$

Proof:

$$f: \quad (a_1, f(a_1))$$

$$g: \quad (a_1, g(a_1))$$

Proving the NFL Theorem (II)

Consider a fixed search algorithm A.

Claim:

$$\forall f, g\colon\ V(T(A, f, |X|)) = V(T(A, g, |X|)) \Rightarrow f = g$$

Proof:

$$\begin{array}{lll} f: & (a_1, f(a_1)), & (a_2, f(a_2)) \\ g: & (a_1, g(a_1) = f(a_1)), & (a_2, g(a_2)) \end{array}$$

Proving the NFL Theorem (II)

Consider a fixed search algorithm A.

Claim:

$$\forall f, g\colon\; V(T(A, f, |X|)) = V(T(A, g, |X|)) \Rightarrow f = g$$

Proof:

$$\begin{array}{lllll} f: & (a_1, f(a_1)), & (a_2, f(a_2)), & & (a_3, f(a_3)), \ldots \\ g: & (a_1, f(a_1)), & (a_2, g(a_2) = f(a_2)), & & (a_3, g(a_3)), \ldots \end{array}$$

$\square$

Proving the NFL Theorem (III)

Consider a fixed search algorithm A.

Claim:

$$\left|\left\{V(T(A, f, |X|)) \mid f \in Y^X\right\}\right| = |Y|^{|X|}$$

Proving the NFL Theorem (III)

Consider a fixed search algorithm A.

Claim:

$$\left|\left\{V(T(A, f, |X|)) \mid f \in Y^X\right\}\right| = |Y|^{|X|}$$

Proof:

$$\left|\left\{f \mid f \in Y^X\right\}\right| = |Y|^{|X|}$$

Proving the NFL Theorem (III)

Consider a fixed search algorithm A.

Claim:

$$\left|\left\{V(T(A, f, |X|)) \mid f \in Y^X\right\}\right| = |Y|^{|X|}$$

Proof:

$$\left|\left\{f \mid f \in Y^X\right\}\right| = |Y|^{|X|}$$

$$\forall f, g \in Y^X :$$
$$f \neq g \Rightarrow V(T(A, f, |X|)) \neq V(T(A, g, |X|))$$

□

Proving the NFL Theorem (IV)

Conclusion:

$$\forall A, B: \quad \{V(T(A, f, |X|)) \mid f \in Y^X\} = \{V(T(B, f, |X|)) \mid f \in Y^X\}$$

Conclusion:

For all finite sets X, Y,
and for all global performance measures,

all deterministic search algorithms perform equal.

Proving the NFL Theorem (V)

X, Y finite $\Rightarrow Y^X$ finite $\Rightarrow$ number of essentially different deterministic search algorithms finite

Proving the NFL Theorem (V)

X, Y finite $\Rightarrow Y^X$ finite $\Rightarrow$ number of essentially different deterministic search algorithms finite

$\Rightarrow$ "randomized search algo." $=$ "prob. dist. over essentially different deterministic search algo."

Proving the NFL Theorem (V)

X, Y finite $\Rightarrow Y^X$ finite $\Rightarrow$ number of essentially different deterministic search algorithms finite

$\Rightarrow$ "randomized search algo." $=$ "prob. dist. over essentially different deterministic search algo."

$\Rightarrow$ "performance of randomized search algorithm" $=$ "weighted sum of the performance of some deterministic search algorithms"

Proving the NFL Theorem (V)

X, Y finite $\Rightarrow Y^X$ finite $\Rightarrow$ number of essentially different deterministic search algorithms finite

$\Rightarrow$ "randomized search algo." $=$ "prob. dist. over essentially different deterministic search algo."

$\Rightarrow$ "performance of randomized search algorithm" $=$ "weighted sum of the performance of some deterministic search algorithms"

$\Rightarrow$ For all finite sets X, Y, and for all global performance measures, all search algorithms perform equal.

Overview

- “The Original” NFL Theorem √
- An improved NFL Theorem
- NFL Assumptions
  - Objective Functions
  - Performance Measures
  - Optimization Scenarios
- NFL Implications

Improved NFL (I)

NFL theorems do not only hold for global measures considering all possible functions.

Improved NFL (I)

NFL theorems do not only hold for global measures considering all possible functions.

Permutations of functions:

Definition: $\sigma \in \text{Perm}(X)\colon \sigma f(x) := f(\sigma^{-1}(x))$

Definition: $\mathcal{F} \subseteq Y^X$ closed under permutations
$\Leftrightarrow \forall f \in \mathcal{F}\colon \forall \sigma \in \text{Perm}(X)\colon \sigma f \in \mathcal{F}$

Definition: performance measure global on $\mathcal{F}$
$\Leftrightarrow \left(\sum_{f \in \mathcal{F}} M(V(A, f, |X|)) \right) / |\mathcal{F}|$

Improved NFL (II)

Improved NFL Theorem:

For all finite sets X, Y,
for all sets $\mathcal{F} \subseteq Y^X$ closed under permutations,
and for all performance measures global on $\mathcal{F}$,
all search algorithms perform equal.

Improved NFL (II)

Improved NFL Theorem:

For all finite sets X, Y,
for all sets $\mathcal{F} \subseteq Y^X$ closed under permutations,
and for all performance measures global on $\mathcal{F}$,
all search algorithms perform equal.

Proof:

$$\forall f \in Y^X : \forall A, B : \exists \sigma \in \text{Perm}(X) :$$
$$V(A, f, |X|) = V(B, \sigma f, |X|)$$

□

Improved NFL (III)

NFL theorems do not hold for performance measures global on arbitrary $\mathcal{F}$.

Improved NFL (III)

NFL theorems do not hold for performance measures global on arbitrary $\mathcal{F}$.

Improved NFL Theorem:

For all finite sets X, Y,
for all sets $\mathcal{F} \subseteq Y^X$,
all search algorithms perform equal
for all performance measures global on $\mathcal{F}$,

if and only if $\mathcal{F}$ is closed under permutations.

Improved NFL (IV)

Proof: by contradiction

X, Y, algorithm A fixed
$\mathcal{F} \subseteq Y^X$ not closed under permutations

$f \in \mathcal{F}$, $\sigma \in$ Perm(X), $\sigma f \notin \mathcal{F}$

Define M: $M(V) = \begin{cases} 1 & \text{if } V = V(T(A, f, |X|)) \\ 0 & \text{otherwise} \end{cases}$

Improved NFL (V)

Obvious: $\sum_{g \in \mathcal{F}} M(V(T(A, g, |X|))) = 1$

We assume NFL holds for $\mathcal{F}$.

Thus for all search algorithms B:

$$\sum_{g \in \mathcal{F}} M(V(T(B, g, |X|))) = 1$$

There is an algorithm C such that

$V(T(C, \sigma f, |X|)) = V(T(A, f, |X|))$.

$\sigma f \notin \mathcal{F} \Rightarrow \sum_{g \in \mathcal{F}} M(V(T(C, g, |X|))) = 0$

contradiction

Overview

- “The Original” NFL Theorem √
- An improved NFL Theorem √
- NFL Assumptions
  - Objective Functions
  - Performance Measures
  - Optimization Scenarios
- NFL Implications

NFL Assumptions — Objective Functions

Original NFL: all functions

Example: Assume $X = \{0,1\}^{40}, Y = \{0,1\}^{8}$.
Assume objective function f is coded in at most 1 giga byte of RAM.

Only $100 \cdot \frac{1024^8 \cdot 8}{(2^8)^{2^{40}}} \% < 10^{-8,796,093,022,120} \%$ of all functions possible.

Most functions can neither be represented nor evaluated.

NFL Assumptions — Objective Functions

Improved NFL:

$\mathcal{F} \subseteq Y^X$ closed under permutations

Claim:

fraction of non-empty $\mathcal{F}$ closed under permutations:

$$\frac{2^{\binom{|X|+|Y|-1}{|X|}} - 1}{2^{|Y|^{|X|}} - 1}$$

Sets with NFL results exponentially small fraction.

NFL Assumptions — Objective Functions

The number of sets closed under permutations:

Definition: For $f : X \rightarrow Y$: histogram $h_f : Y \rightarrow \mathbb{N}_0$ with $h_f(y) = |\{x \in X \mid f(x) = y\}|$

Observation: "$f \sim g \Leftrightarrow h_f = h_g$" defines equivalence relation

Definition: basis class B_f: equivalence class of f

NFL Assumptions — Objective Functions

The number of sets closed under permutations:

Claim: There are $\binom{|X|+|Y|-1}{|X|}$ pairwise disjoint basis classes.

Proof: basis class "$\triangleq$" histogram

$$\left| \begin{array}{c} 1 \\ n_1 \end{array} \right| \begin{array}{c} 2 \\ n_2 \end{array} \left| \begin{array}{c} 3 \\ n_3 \end{array} \right| \begin{array}{c} \ldots \\ \ldots \end{array} \left| \begin{array}{c} |Y|-1 \\ n_{|Y|-1} \end{array} \right| \begin{array}{c} |Y| \\ n_{|Y|} \end{array} \Bigg| \text{ with } \sum_{i=1}^{|Y|} n_i = |X|$$

Think of each image is a ball:
Pick pos. of $|X|$ balls from $|X| + |Y| - 1$ pos.

NFL Assumptions — Objective Functions

The number of sets closed under permutations:

Claim: $f \sim g \Leftrightarrow \exists \sigma \in \text{Perm}(X) : \sigma f = g$

Proof:

"$\Rightarrow$": Define σ.

"$\Leftarrow$": histogram permutation-invariant

$\square$

NFL Assumptions — Objective Functions

The number of sets closed under permutations:

Claim: $\forall \mathcal{F} \subseteq Y^X$ closed under permutation:
$\mathcal{F}$ is union of basis classes

Proof:
For $f \in \mathcal{F}$ consider $F_f := B_f \cap \mathcal{F}$.

Observations:

$$\bigcup_{f \in \mathcal{F}} F_f = \mathcal{F} \qquad F_f \subseteq B_f \qquad B_f \subseteq F_f$$

□

NFL Assumptions — Objective Functions

The number of sets closed under permutations:

Claim:
fraction of non-empty $\mathcal{F}$ closed under permutations:

$$\frac{2^{\binom{|X|+|Y|-1}{|X|}} - 1}{2^{|Y|^{|X|}} - 1}$$

Proof: number of non-empty unions of basis classes: $2^{\text{number of basis classes}} - 1$
number of all non-empty sets: $2^{\text{number of functions}} - 1$

□

NFL Assumptions — Objective Functions

Definition: neighborhood $N\colon X \times X \to \{\text{true}, \text{false}\}$

Definition: neighborhood N non-trivial $\Leftrightarrow$
$\exists x_1, x_2, x_3, x_4 \in X :$
$(x_1 \neq x_2 \wedge N(x_1, x_2) = \text{true}) \wedge$
$(x_3 \neq x_4 \wedge N(x_3, x_4) = \text{false})$

Observation:
Any non-trivial neighborhood is not invariant under permutations.

NFL Assumptions — Objective Functions

Let $\mathcal{F} \subseteq Y^X$ be the set of “possible functions”.

NFL Assumptions — Objective Functions

Let $\mathcal{F} \subseteq Y^X$ be the set of “possible functions”.

Conclusion:

For $\mathcal{F}$ with minimal “smoothness”
NFL results do not hold.

NFL Assumptions — Objective Functions

Let $\mathcal{F} \subseteq Y^X$ be the set of “possible functions”.

Conclusion:

For $\mathcal{F}$ with minimal “smoothness”
NFL results do not hold.

Conclusion:

For $\mathcal{F}$ with a bounded number of local optima
NFL results do not hold.

NFL Assumptions — Performance Measures

Randomized search heuristics do re-evaluate.

Re-evaluations do influence the computation time.

NFL results have limited implications for consideration of computation time.

NFL Assumptions — Optimization Scenarios

Black Box Optimization:

Randomized search heuristic has to work for all $f \in \mathcal{F}$, where $\mathcal{F} \subseteq Y^X$ is complexity restricted.

Possible complexity restrictions:

- evaluation time restricted
- representation size restricted
- Kolmogoroff complexity restricted

The ANFL Theorem

The Almost No Free Lunch Theorem:

$X = \{0,1\}^n$, $Y = \{0,1,\dots,N-1\}$, $f\colon X \to Y$,
A a randomized search algorithm

$|\{f^*\colon \{0,1\}^n \to \{0,1,\dots,N\} \mid f^*$ agrees with f on $\geq 2^n - 2^{n/3}$ inputs and A finds the max. of f^* within $2^{n/3}$ steps with prob. $\leq 2^{-n/3}\ \}| \geq N^{2^{n/3}-1}$.

For exp. many functions $f^*\colon C(f^*) = C(f) + O(n)$

where C is evaluation time, circuit representation size, or Kolmogoroff complexity

Proof of the ANFL Theorem (I)

A: randomized search algorithm

$f:\{0,1\}^n \to \{0,1,\ldots,N-1\}$: function,
such that A is efficient on f

Consider first $2^{n/3}$ steps of A on f.

$q(x) = \text{Prob}(A \text{ visits } x \text{ in the first } 2^{n/3} \text{ steps})$

$$\sum_{x\in\{0,1\}^n} q(x)$$

Proof of the ANFL Theorem (I)

A: randomized search algorithm

$f: \{0,1\}^n \to \{0,1,\ldots,N-1\}$: function,
such that A is efficient on f

Consider first $2^{n/3}$ steps of A on f.

$q(x) =$ Prob(A visits x in the first $2^{n/3}$ steps)

$$\sum_{x \in \{0,1\}^n} q(x) \leq 2^{n/3}$$

Proof of the ANFL Theorem (II)

Definition for $b \in \{0,1\}^{2n/3}$:

$$S_b := \{x \in \{0,1\}^n \mid \forall i \in \{1,2,\ldots,2n/3\}\colon x_i = b_i\}$$

Observations:

- $\forall b \in \{0,1\}^{2n/3}\colon\ |S_b| = 2^{n/3}$
- $\left|\left\{S_b \mid b \in \{0,1\}^{2n/3}\right\}\right| = 2^{2n/3}$
- $\forall b \neq b' \in \{0,1\}^{2n/3}\colon S_b \cap S_{b'} = \emptyset$

Proof of the ANFL Theorem (III)

Definition:

$q^*(b) :=$ Prob(A visits S_b in the first $2^{n/3}$ steps)

Proof of the ANFL Theorem (III)

Definition:

$q^*(b) :=$ Prob(A visits S_b in the first $2^{n/3}$ steps)

Obvious: $q^*(b) \leq \sum_{x \in S_b} q(x)$

Proof of the ANFL Theorem (III)

Definition:

$q^*(b) :=$ Prob$($$A$ visits S_b in the first $2^{n/3}$ steps$)$

Obvious: $q^*(b) \leq \sum_{x \in S_b} q(x)$

Conclusion: $\exists b^*: q^*(b^*) \leq \frac{2^{n/3}}{2^{2n/3}} = 2^{-n/3}$

Proof of the ANFL Theorem (III)

Definition:

$f^*: \{0,1\}^n \longrightarrow \mathbb{R}$ with

- $\forall x \notin S_{b^*}: f^*(x) = f(x)$
- $\exists x^* \in S_{b^*}: f^*(x^*) = N$

Proof of the ANFL Theorem (III)

Definition:

$f^*: \{0,1\}^n \to \mathbb{R}$ with

- $\forall x \notin S_{b^*}: f^*(x) = f(x)$
- $\exists x^* \in S_{b^*}: f^*(x^*) = N$

Observations:

- There are at least $N^{2^{n/3}} - 1$ such f^*.
- Prob(A optimizes f^* in $\leq 2^{n/3}$ steps) $\leq 2^{-n/3}$

Proof of the ANFL Theorem (III)

Problem: Almost all such f^* are hard to represent and evaluate.

Solution:

$$f^*_{x^*,c}(x) = \begin{cases} f(x) & x \notin S_{b^*} \\ N & x = x^* \\ c & \text{otherwise} \end{cases}$$

There are $N \cdot 2^{n/3}$ such functions $f^*_{x^*,c}$. $\square$

Overview

- “The Original” NFL Theorem √
- An improved NFL Theorem √
- NFL Assumptions √
  - Objective Functions √
  - Performance Measures √
  - Optimization Scenarios √
- NFL Implications

NFL Implications

- The (improved) NFL is a valid theorem.

NFL Implications

- The (improved) NFL is a valid theorem.
- Be careful with claims like "algorithm A is better than algorithm B".

NFL Implications

- The (improved) NFL is a valid theorem.
- Be careful with claims like “algorithm A is better than algorithm B”.
- If someone claims “In my application domain NFL results do not hold” he or she is prob. right.

NFL Implications

- The (improved) NFL is a valid theorem.
- Be careful with claims like “algorithm A is better than algorithm B”.
- If someone claims “In my application domain NFL results do not hold” he or she is prob. right.
- When measuring computation time, NFL is not helpful.

NFL Implications

- The (improved) NFL is a valid theorem.
- Be careful with claims like “algorithm A is better than algorithm B”.
- If someone claims “In my application domain NFL results do not hold” he or she is prob. right.
- When measuring computation time, NFL is not helpful.
- NFL has little impact on research.

NFL Implications

- The (improved) NFL is a valid theorem.
- Be careful with claims like “algorithm A is better than algorithm B”.
- If someone claims “In my application domain NFL results do not hold” he or she is prob. right.
- When measuring computation time, NFL is not helpful.
- NFL has little impact on research.
- The ANFL Theorem restricts what can be achieved.

Overview

- “The Original” NFL Theorem √
- An improved NFL Theorem √
- NFL Assumptions √
  - Objective Functions √
  - Performance Measures √
  - Optimization Scenarios √
- NFL Implications √