The Sample Complexity of Level Set Approximation

François Bachoc[1,2] Tom Cesari[3,4] Sébastien Gerchinovitz[5,2]

[1]UT3 [2]IMT [3]TSE [4]ANITI [5]IRT

EUROPT 2021

Problem, motivations, related work

Problem

Approximating $\{\boldsymbol{x} : f(\boldsymbol{x}) = a\} \subset [0,1]^d$

- $f : [0,1]^d \to \mathbb{R}$ unknown in some known smoothness class
- $a \in \mathbb{R}$ a fixed known threshold

Problem, motivations, related work

Problem

Approximating $\{\boldsymbol{x} : f(\boldsymbol{x}) = a\} \subset [0,1]^d$

- $f : [0,1]^d \to \mathbb{R}$ unknown in some known smoothness class
- $a \in \mathbb{R}$ a fixed known threshold

Motivation

Determining parameters that result in a given outcome (computer experiments, uncertainty quantification, nuclear engineering, coastal flooding, etc.)

Problem, motivations, related work

Problem

Approximating $\{\boldsymbol{x} : f(\boldsymbol{x}) = a\} \subset [0, 1]^d$

- $f : [0, 1]^d \to \mathbb{R}$ unknown in some known smoothness class
- $a \in \mathbb{R}$ a fixed known threshold

Motivation

Determining parameters that result in a given outcome (computer experiments, uncertainty quantification, nuclear engineering, coastal flooding, etc.)

Related work

- Gaussian process models: Chevalier et al. 2014 Technometrics, Azzimonti et al. 2020 Technometrics, Gotovos et al. 2013 IJCAI
- Global optimization algorithms: DOO, Munos 2011 NIPS, HOO, Bubeck et al. 2011 JMLR

Protocol and Objective

Online Protocol

For $n = 1, 2, \ldots$

1. pick the next query point $\boldsymbol{x}_n$
2. observe the value $f(\boldsymbol{x}_n)$
3. output an approximating set S_n

Protocol and Objective

Online Protocol

For $n = 1, 2, \ldots$

1. pick the next query point $\boldsymbol{x}_n$
2. observe the value $f(\boldsymbol{x}_n)$
3. output an approximating set S_n

Our Goal

Quantifying the sample complexity, i.e., smallest number of evaluations of f needed to

$$\{\boldsymbol{x} : f(\boldsymbol{x}) = a\} \subset S_n \subset \{\boldsymbol{x} : |f(\boldsymbol{x}) - a| \leq \varepsilon\}$$

for some error level $\varepsilon > 0$

A Hard Problem

Definition

The packing number of a non-empty set E is

$$\mathcal{N}(E,\varepsilon) := \sup\Big\{k \in \mathbb{N} : \exists\, \boldsymbol{x}_1,\ldots,\boldsymbol{x}_k \in E,\ \min_{i\neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_\infty > \varepsilon\Big\}$$

A Hard Problem

Definition

The packing number of a non-empty set E is

$$\mathcal{N}(E,\varepsilon) := \sup\Big\{k \in \mathbb{N} : \exists\, \boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in E,\, \min_{i\neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_\infty > \varepsilon\Big\}$$

Theorem

If f is a non-constant continuous function, for any $\min(f) < a < \max(f)$,

$$\mathcal{N}\Big(\{f = a\}, \varepsilon\Big) \gtrsim \frac{1}{\varepsilon^{d-1}}$$

A Hard Problem

Definition

The packing number of a non-empty set E is

$$\mathcal{N}(E,\varepsilon) := \sup\Big\{k \in \mathbb{N} : \exists\, \boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in E,\ \min_{i\neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_\infty > \varepsilon\Big\}$$

Theorem

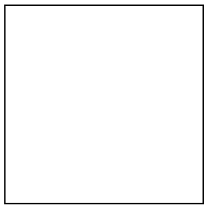
If f is a non-constant continuous function, for any $\min(f) < a < \max(f)$,

$$\mathcal{N}\Big(\{f = a\},\varepsilon\Big) \gtrsim \frac{1}{\varepsilon^{d-1}}$$

Not surprising, the level set is defined by a single equation in d unknowns

A General Solution

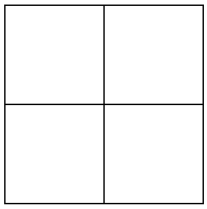
Bisect and Approximate (BA)



A General Solution

Bisect and Approximate (BA)

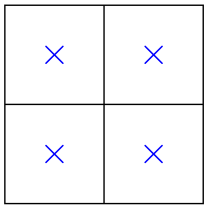
1. Bisect the current family of cells



A General Solution

Bisect and Approximate (BA)

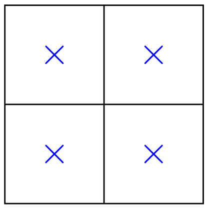
1. Bisect the current family of cells
2. Query f at some point(s) in each new cell



A General Solution

Bisect and Approximate (BA)

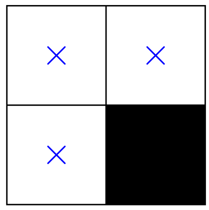
1. Bisect the current family of cells
2. Query f at some point(s) in each new cell
3. Compute a local approximator g_C of f on each cell C



A General Solution

Bisect and Approximate (BA)

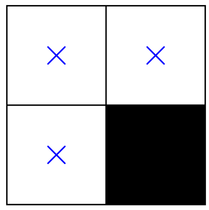
1. Bisect the current family of cells
2. Query f at some point(s) in each new cell
3. Compute a local approximator g_C of f on each cell C
4. Remove a cell C if $|g_C(\boldsymbol{x}) - a|$ is large for all $\boldsymbol{x} \in C$



A General Solution

Bisect and Approximate (BA)

1. Bisect the current family of cells
2. Query f at some point(s) in each new cell
3. Compute a local approximator g_C of f on each cell C
4. Remove a cell C if $|g_C(\boldsymbol{x}) - a|$ is large for all $\boldsymbol{x} \in C$



Theorem

If the g_C's are "accurate approximations" of f on the C's,

$$\text{sample complexity of BA} \lesssim \sum_{i=1}^{i(\varepsilon)} \mathcal{N}\Big(\{|f-a| \le c_i\}, d_i\Big)$$

where $i(\varepsilon) \approx \log(1/\varepsilon)$, $c_1 > c_2 > \dots$, $d_1 > d_2 > \dots$ depend on the g_C's and their error bounds

Consequence for γ-Hölder functions

γ-Hölder functions

f is γ-Hölder ($|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq c \, \|\boldsymbol{x} - \boldsymbol{y}\|^{\gamma}$, with $\gamma \in (0, 1]$)

Consequence for γ-Hölder functions

γ-Hölder functions

f is γ-Hölder ($|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq c \, \|\boldsymbol{x} - \boldsymbol{y}\|^{\gamma}$, with $\gamma \in (0, 1]$)

Local approximator for BAH

- For a cell C (hypercube), we query the center and take the local approximator g_C as constant
- The error of g_C on C is $\lesssim \mathrm{Diam}(C)^{\gamma}$

Consequence for γ-Hölder functions

γ-Hölder functions

f is γ-Hölder ($|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq c \, \|\boldsymbol{x} - \boldsymbol{y}\|^\gamma$, with $\gamma \in (0, 1]$)

Local approximator for BAH

- For a cell C (hypercube), we query the center and take the local approximator g_C as constant
- The error of g_C on C is $\lesssim \mathrm{Diam}(C)^\gamma$

Theorem (upper and lower bound)

The worst-case optimal sample complexity is attained by BAH and

$$\text{sample complexity of BAH} \lesssim \frac{1}{\varepsilon^{d/\gamma}}$$

For lower bound counter example functions are “flat + local bump” (classical)

Consequence for functions with γ_1-Hölder gradients

Functions with γ_1-Hölder gradients

∇f is γ_1-Hölder

Consequence for functions with γ_1-Hölder gradients

Functions with γ_1-Hölder gradients

∇f is γ_1-Hölder

Local approximator for BAG

- For a cell C (hypercube), we query the 2^d vertices and take the local approximator g_C as multilinear interpolating
- The error of g_C on C is $\lesssim \mathrm{Diam}(C)^{1+\gamma_1}$

Consequence for functions with γ_1-Hölder gradients

Functions with γ_1-Hölder gradients

∇f is γ_1-Hölder

Local approximator for BAG

- For a cell C (hypercube), we query the 2^d vertices and take the local approximator g_C as multilinear interpolating
- The error of g_C on C is $\lesssim \mathrm{Diam}(C)^{1+\gamma_1}$

Theorem (upper and lower bound)

The worst-case optimal sample complexity is attained by BAG and

$$\text{sample complexity of BAG} \lesssim \frac{1}{\varepsilon^{d/(1+\gamma_1)}}$$

For lower bound counter example functions are “flat + local bump” (classical)

Convexity helps a little: $d \mapsto d - 1$

Theorem

When f has γ_1-Hölder gradient and is convex (+ quantitative conditions), then the worst-case optimal sample complexity is attained by BAG and

$$\text{sample complexity of BAG} \lesssim \frac{1}{\varepsilon^{(d-1)/(1+\gamma_1)}}$$

Convexity helps a little: $d \mapsto d-1$

Theorem

When f has γ_1-Hölder gradient and is convex (+ quantitative conditions), then the worst-case optimal sample complexity is attained by BAG and

$$\text{sample complexity of BAG} \lesssim \frac{1}{\varepsilon^{(d-1)/(1+\gamma_1)}}$$

Follows from geometric arguments on level sets of convex functions:

Theorem

If f is convex (+ quantitative conditions), there exists a constant $C^* > 0$ such that

$$\forall r \in (0,1)\ ,\ \mathcal{N}\Big(\{|f-a| \le r\},\ r\Big) \le C^* \left(\frac{1}{r}\right)^{d-1}$$

Open Problems

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$,

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$, $\partial^{(1)} f$ γ_1-Hölder $\Rightarrow 1/\varepsilon^{d/(1+\gamma_1)}$,

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$, $\partial^{(1)} f$ γ_1-Hölder $\Rightarrow 1/\varepsilon^{d/(1+\gamma_1)}$, $\partial^{(k)} f$ γ_k-Hölder $\Rightarrow 1/\varepsilon^{d/(k+\gamma_k)}$ (similar results in nonparametric statistics theory)

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$, $\partial^{(1)} f$ γ_1-Hölder $\Rightarrow 1/\varepsilon^{d/(1+\gamma_1)}$, $\partial^{(k)} f$ γ_k-Hölder $\Rightarrow 1/\varepsilon^{d/(k+\gamma_k)}$ (similar results in nonparametric statistics theory)
- Adaptivity to smoothness

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$, $\partial^{(1)} f$ γ_1-Hölder $\Rightarrow 1/\varepsilon^{d/(1+\gamma_1)}$, $\partial^{(k)} f$ γ_k-Hölder $\Rightarrow 1/\varepsilon^{d/(k+\gamma_k)}$ (similar results in nonparametric statistics theory)
- Adaptivity to smoothness

The paper:

F. Bachoc, T. Cesari and S. Gerchinovitz, "The sample complexity of level set approximation" AISTATS 2021 - oral presentation

Open Problems

- f γ-Hölder $\Rightarrow 1/\varepsilon^{d/\gamma}$, $\partial^{(1)} f$ γ_1-Hölder $\Rightarrow 1/\varepsilon^{d/(1+\gamma_1)}$, $\partial^{(k)} f$ γ_k-Hölder $\Rightarrow 1/\varepsilon^{d/(k+\gamma_k)}$ (similar results in nonparametric statistics theory)
- Adaptivity to smoothness

The paper:

F. Bachoc, T. Cesari and S. Gerchinovitz, "The sample complexity of level set approximation" AISTATS 2021 - oral presentation