ZEROTH-ORDER NON-CONVEX SMOOTH OPTIMIZATION: LOCAL MINIMAX RATES

Yining Wang, CMU

joint work with Sivaraman Balakrishnan and Aarti Singh

BACKGROUND

- Optimization: $\min_{x \in \mathcal{X}} f(x)$
- Classical setting (first-order):
  - f is known (e.g., a likelihood function or an NN objective)
  - $\nabla f(x)$ can be evaluated, or unbiasedly approximated.
- Zeroth-order setting:
  - f is unknown, or very complicated.
  - $\nabla f(x)$ is unknown, or very difficult to evaluate.
  - $f(x)$ can be evaluated, or unbiasedly approximated.

BACKGROUND

- Hyper-parameter tuning
  - *f* maps hyper-parameter θ to system performance *r*
  - *f* is essentially unknown
- Experimental design
  - *f* maps experimental setting (pressure, temperature, etc.) to synthesized material quality.
- Communication-efficient optimization
  - Data defining the objective scattered throughout machines
  - Communicating $\nabla f(x)$ is expensive, but $f(x)$ is ok.

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α $\quad \|f^{(\alpha)}\|_\infty \leq M$
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α $\longrightarrow$ $\|f^{(\alpha)}\|_\infty \leq M$
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α $\longrightarrow$ $\|f^{(\alpha)}\|_\infty \leq M$
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$

PROBLEM FORMULATION

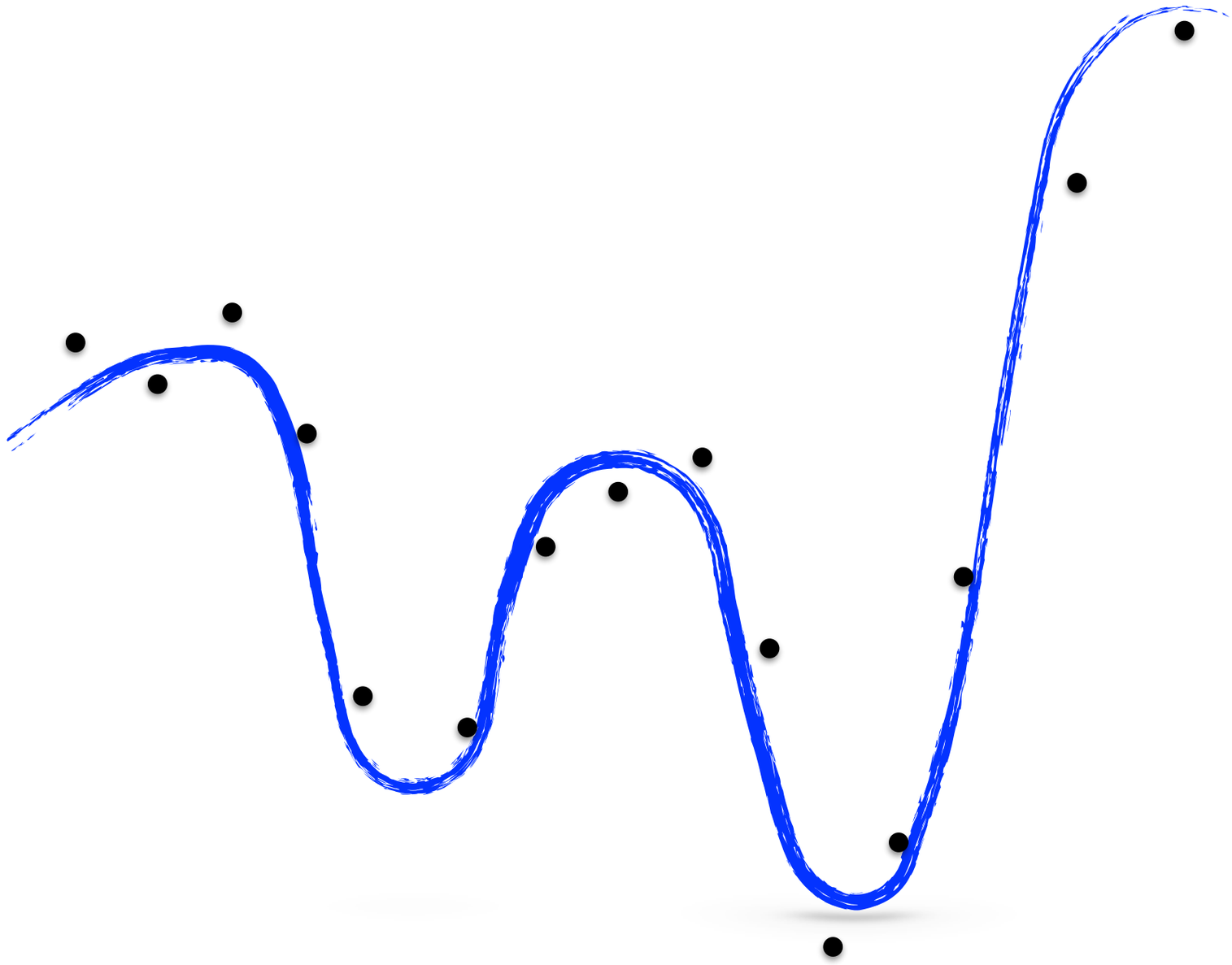
- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α $\longrightarrow$ $\|f^{(\alpha)}\|_\infty \leq M$
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$ $=: \mathfrak{L}(\hat{x}_n; f)$

PROBLEM FORMULATION

- Compact domain $\mathcal{X} = [0,1]^d$
- Objective function $f : \mathcal{X} \to \mathbb{R}$
  - f belongs to the Holder class of order α $\longrightarrow$ $\|f^{(\alpha)}\|_\infty \leq M$
  - f may be non-convex
- Query model: adaptive $x_1, x_2, \cdots, x_n \in \mathcal{X}$
  - $y_t = f(x_t) + \xi_t \quad \xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$
- **Goal: minimize** $f(\hat{x}_n) - \inf_{x \in \mathcal{X}} f(x)$ $\longrightarrow$ $=: \mathfrak{L}(\hat{x}_n; f)$

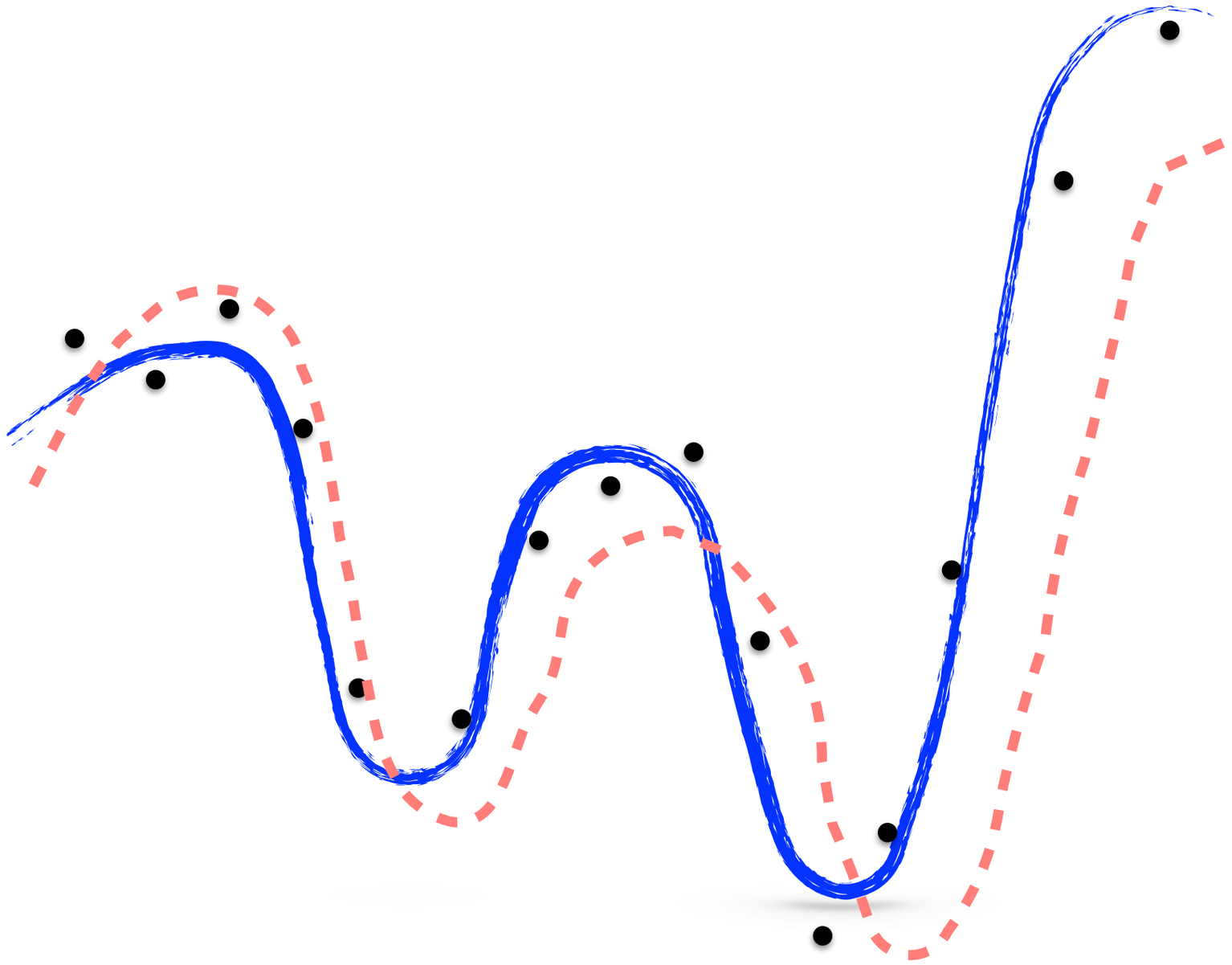
A SIMPLE IDEA FIRST . . .

➤ Uniform sampling + nonparametric reconstruction



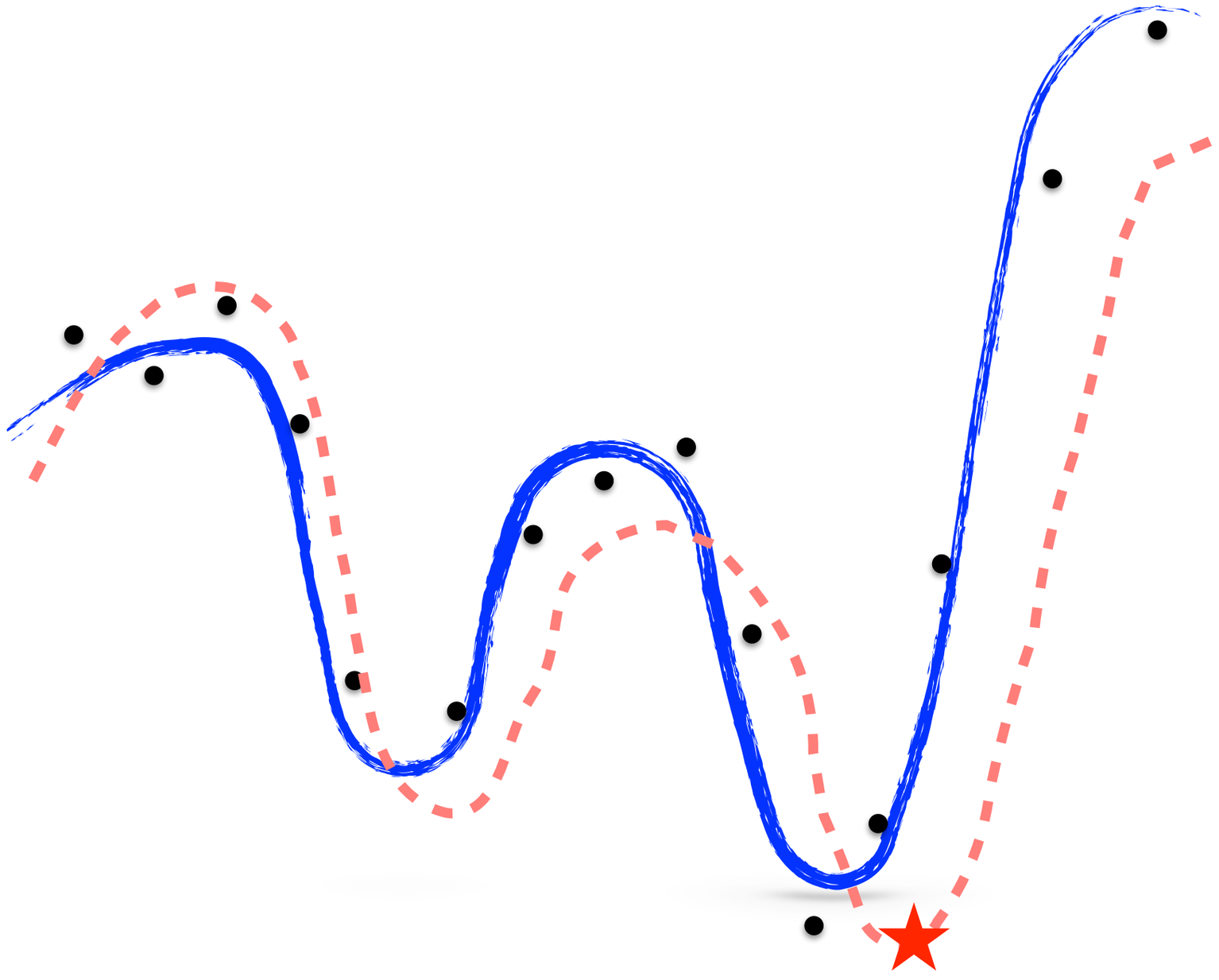
A SIMPLE IDEA FIRST . . .

➤ Uniform sampling + nonparametric reconstruction



A SIMPLE IDEA FIRST . . .

➤ Uniform sampling + nonparametric reconstruction



A SIMPLE IDEA FIRST . . .

- Uniform sampling + nonparametric reconstruction
  - Classical Non-parametric analysis

$$\|\hat{f}_n - f\|_\infty = \widetilde{O}_P\left(n^{-\alpha/(2\alpha+d)}\right)$$

  - Implies optimization error: $f(\hat{x}_n) - f^* \leq 2\|f_n - f\|_\infty$
- Can we do better?
- **NO!**

$$\inf_{\hat{x}_n} \sup_{f \in \Sigma^\alpha(M)} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n; f)\right] \gtrsim n^{-\alpha/(2\alpha+d)}$$

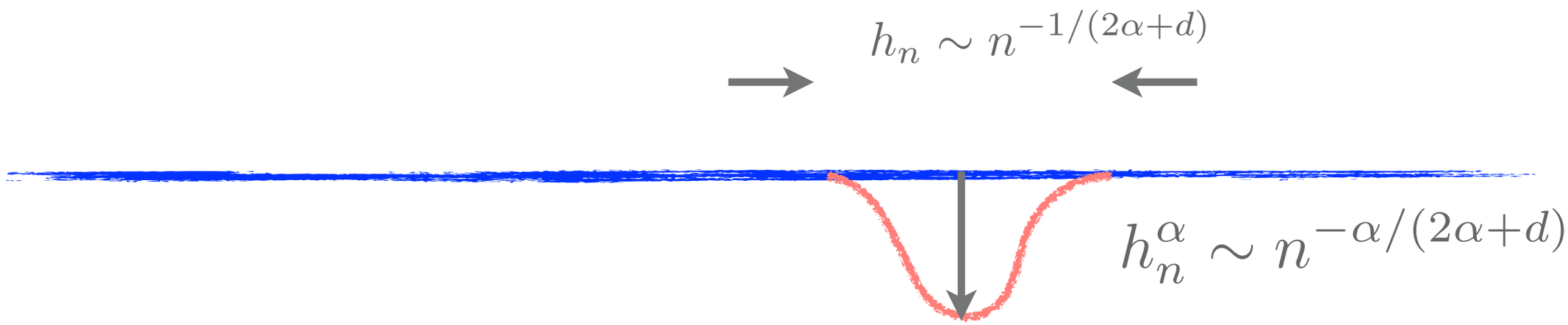
A SIMPLE IDEA FIRST . . .

- Uniform sampling + nonparametric reconstruction
  - Classical Non-parametric analysis

$$\|\hat{f}_n - f\|_\infty = \widetilde{O}_P\left(n^{-\alpha/(2\alpha+d)}\right)$$

  - Implies optimization error: $f(\hat{x}_n) - f^* \leq 2\|f_n - f\|_\infty$

- Can we do better? No! Intuitions:

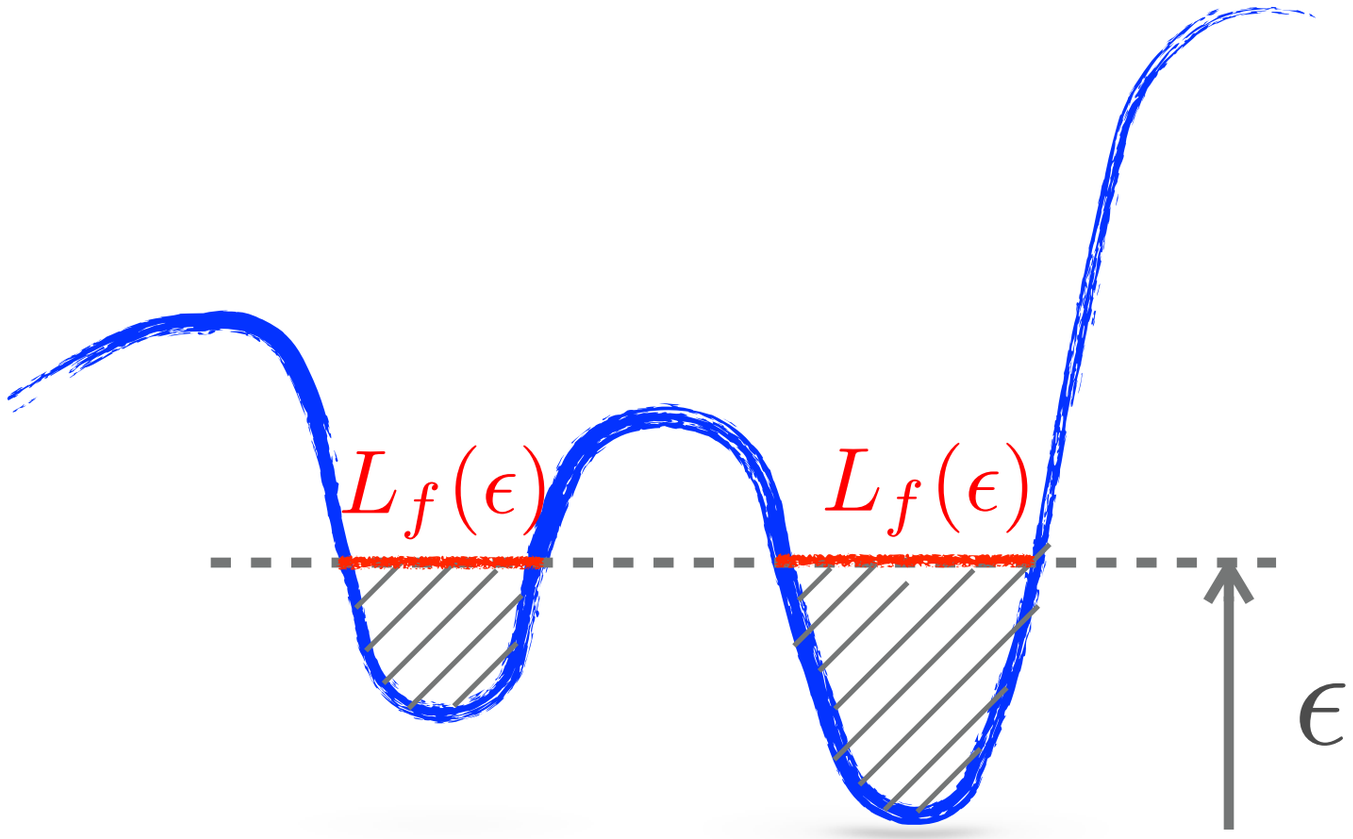


LOCAL RESULTS

- Characterize error for functions "near" a reference function f_0
- What is the error rate for f close to f_0 that is …
  - a constant function?
  - strongly convex?
  - has regular level sets?
  - …
- Can an algorithm achieve instance-optimal error, without knowing f_0?

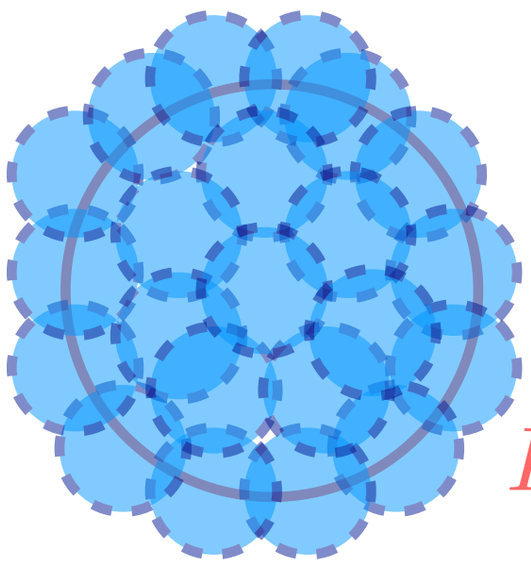
NOTATIONS

- Some definitions
  - Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$



REGULARITY CONDITIONS

- Some definitions
  - Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$
- Regularity condition (A1):
  - # of δ-radius balls needed to cover $L_f(\epsilon) \asymp 1 + \mu_f(\epsilon)/\delta^d$

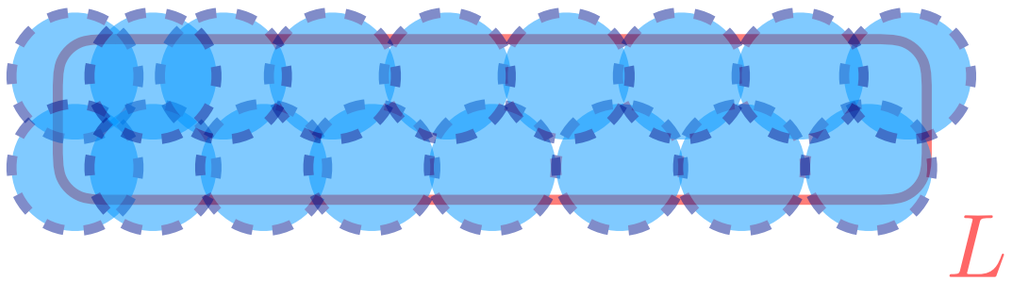


$L_f(\epsilon)$

Regular level-set

REGULARITY CONDITIONS

- Some definitions
  - ✳ Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - ✳ Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$
- Regularity condition (A1):
  - ✳ # of δ-radius balls needed to cover $L_f(\epsilon) \asymp 1 + \mu_f(\epsilon)/\delta^d$

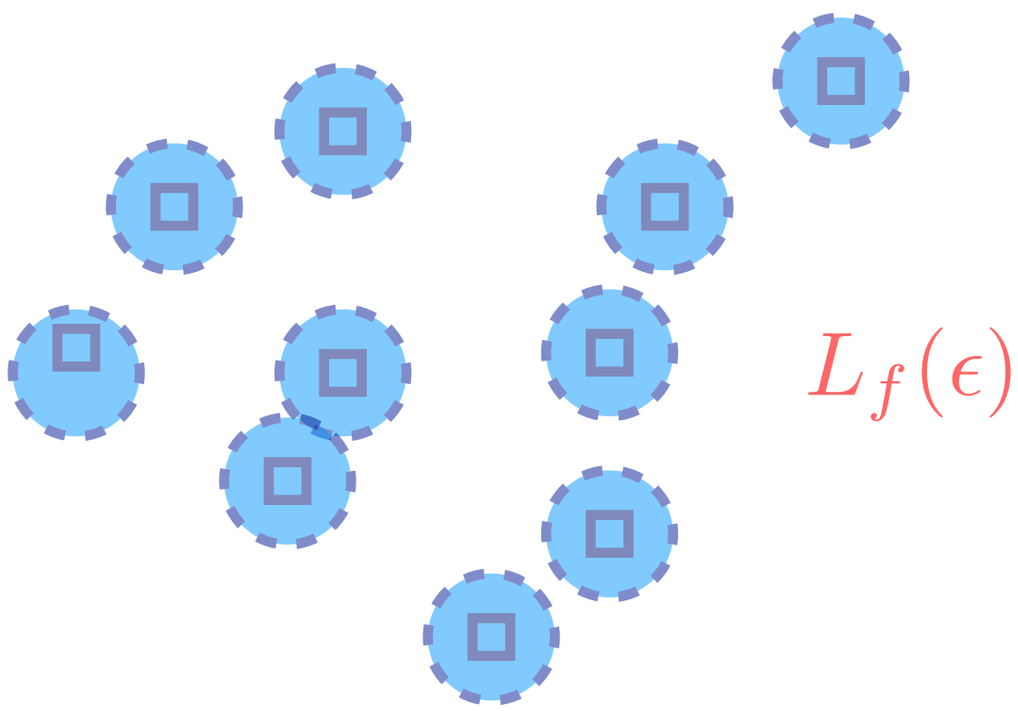


$L_f(\epsilon)$

Irregular level-set

REGULARITY CONDITIONS

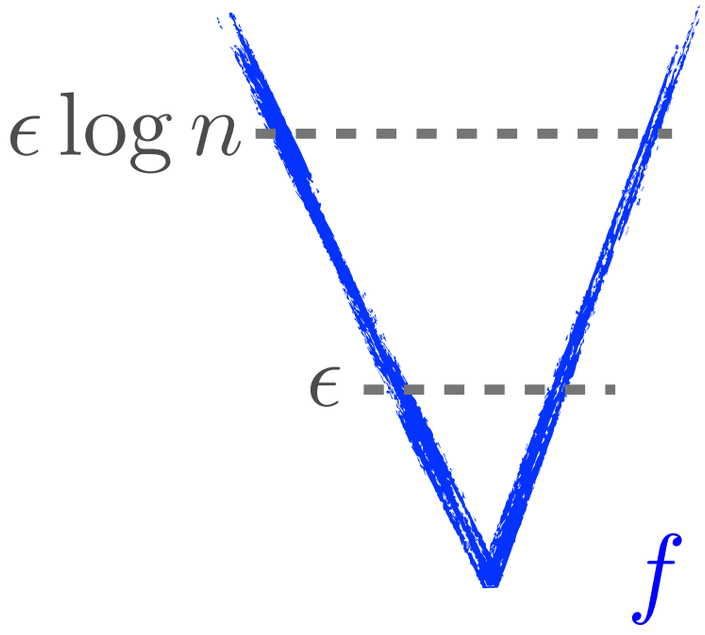
- Some definitions
  - ✳ Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - ✳ Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$
- Regularity condition (A1):
  - ✳ # of δ-radius balls needed to cover $L_f(\epsilon) \asymp 1 + \mu_f(\epsilon)/\delta^d$

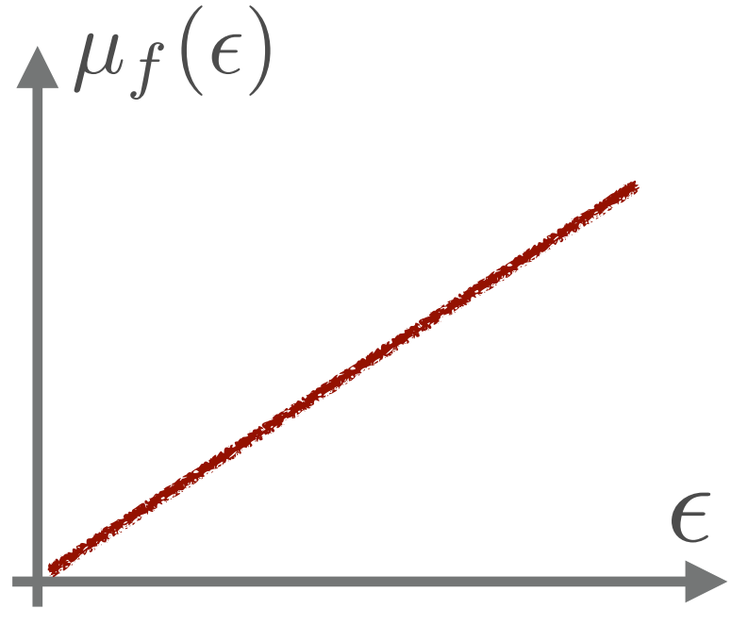


Irregular level-set

REGULARITY CONDITIONS

- Some definitions
  - ✳ Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - ✳ Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$
- Regularity condition (A2):
  - ✳ $\mu_f(\epsilon \log n) \leq \mu_f(\epsilon) \times O(\log^\gamma n)$

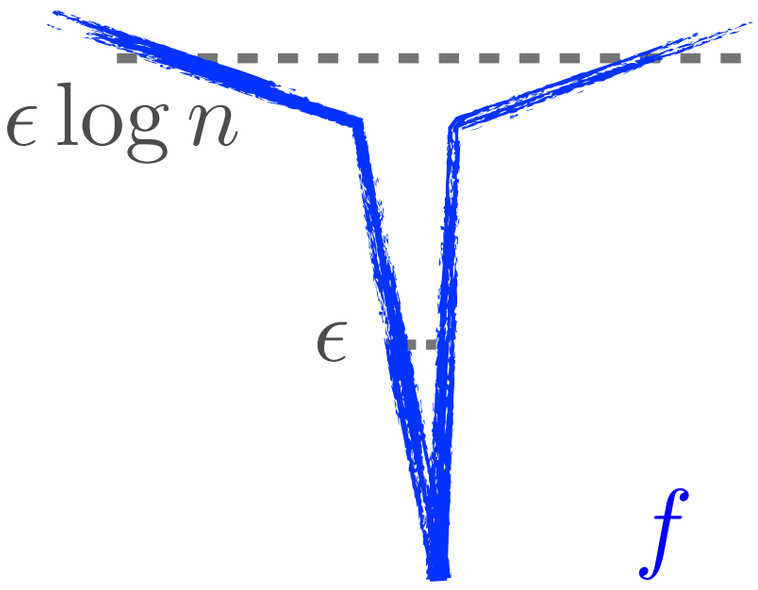


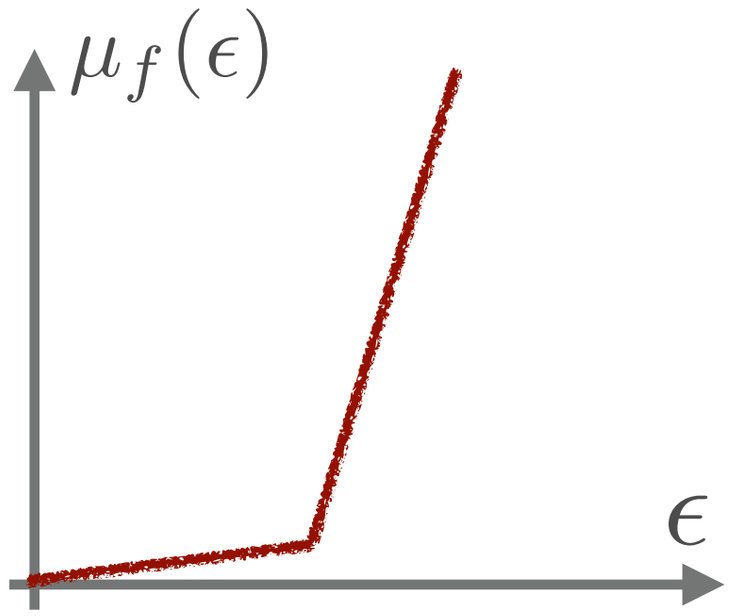


Regular

REGULARITY CONDITIONS

- Some definitions
  - Level set: $L_f(\epsilon) := \{x \in \mathcal{X} : f(x) \leq f^* + \epsilon\}$
  - Distribution function: $\mu_f(\epsilon) := \mathrm{vol}(L_f(\epsilon))$
- Regularity condition (A2):
  - $\mu_f(\epsilon \log n) \leq \mu_f(\epsilon) \times O(\log^\gamma n)$





Irregular

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $$\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$$

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $$\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$$

Adaptivity:

The algo does not know f.

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $$\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$$

Adaptivity:

The algo does not know f.

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $$\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$$

Adaptivity:

The algo does not know f.

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

Adaptivity:

The algo does not know f.

Instance dependent:

Error rate depends on f

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

➤ Example 1: polynomial growth $\mu_f(\epsilon) \asymp \epsilon^\beta, \beta \geq 0$

$$\varepsilon_n(f) \asymp n^{-\alpha/(2\alpha+d-\alpha\beta)}$$

Much faster than the "baseline" rate $n^{-\alpha/(2\alpha+d)}$

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

➤ Example 2: constant function $f \equiv 0$

$$\varepsilon_n(f) \asymp n^{-\alpha/(2\alpha+d)}$$

This is the worst case function, matching global rate $n^{-\alpha/(2\alpha+d)}$

LOCAL UPPER BOUND

➤ Main result on local upper bound:

THEOREM 1. **Suppose regularity conditions hold. There exists an algorithm such that for sufficiently large n,**

$$\sup_{f \in \Sigma^\alpha(M)} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geq C \varepsilon_n(f) \log^c n \right] \leq 1/4$$

where $\varepsilon_n(f) := \sup \left\{ \epsilon > 0 : \epsilon^{-(2+d/\alpha)} \mu_f(\epsilon) \geq n \right\}$

➤ Example 3: strongly convex f: $\mu_f(\epsilon) \asymp \epsilon^{d/2}$

$$\varepsilon_n(f) \asymp n^{-1/2}$$

Match the classical zeroth-order convex rate $n^{-1/2}$ up to log terms

LOCAL LOWER BOUND

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n;f)\right] \gtrsim \varepsilon_n(f_0)$$

where $$\varepsilon_n(f) := \sup\left\{\epsilon>0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon)\geq n\right\}$$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n;f)\right] \gtrsim \varepsilon_n(f_0)$$

where $\varepsilon_n(f) := \sup\left\{\epsilon>0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon)\geq n\right\}$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n;f)\right] \gtrsim \varepsilon_n(f_0)$$

where $$\varepsilon_n(f) := \sup\left\{\epsilon>0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon)\geq n\right\}$$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n;f)\right] \gtrsim \varepsilon_n(f_0)$$

where $$\varepsilon_n(f) := \sup\left\{\epsilon>0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon)\geq n\right\}$$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n;f)\right] \gtrsim \varepsilon_n(f_0)$$

where $$\varepsilon_n(f) := \sup\left\{\epsilon>0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon)\geq n\right\}$$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n; f)\right] \gtrsim \varepsilon_n(f_0)$$

where $\varepsilon_n(f) := \sup\left\{\epsilon > 0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon) \geq n\right\}$

LOCAL LOWER BOUND

full knowledge of reference:

The algo. knows f_0

➤ Main result on local lower bound:

THEOREM 2. **Suppose f_0 satisfies regularity conditions. Then we have**

$$\inf_{\hat{x}_n} \sup_{\substack{f\in\Sigma^\alpha(M),\\ \|f-f_0\|_\infty\lesssim\varepsilon_n(f_0)}} \mathbb{E}_f\left[\mathfrak{L}(\hat{x}_n; f)\right] \gtrsim \varepsilon_n(f_0)$$

where $\varepsilon_n(f) := \sup\left\{\epsilon > 0 : \epsilon^{-(2+d/\alpha)}\mu_f(\epsilon) \geq n\right\}$

Local minimality:

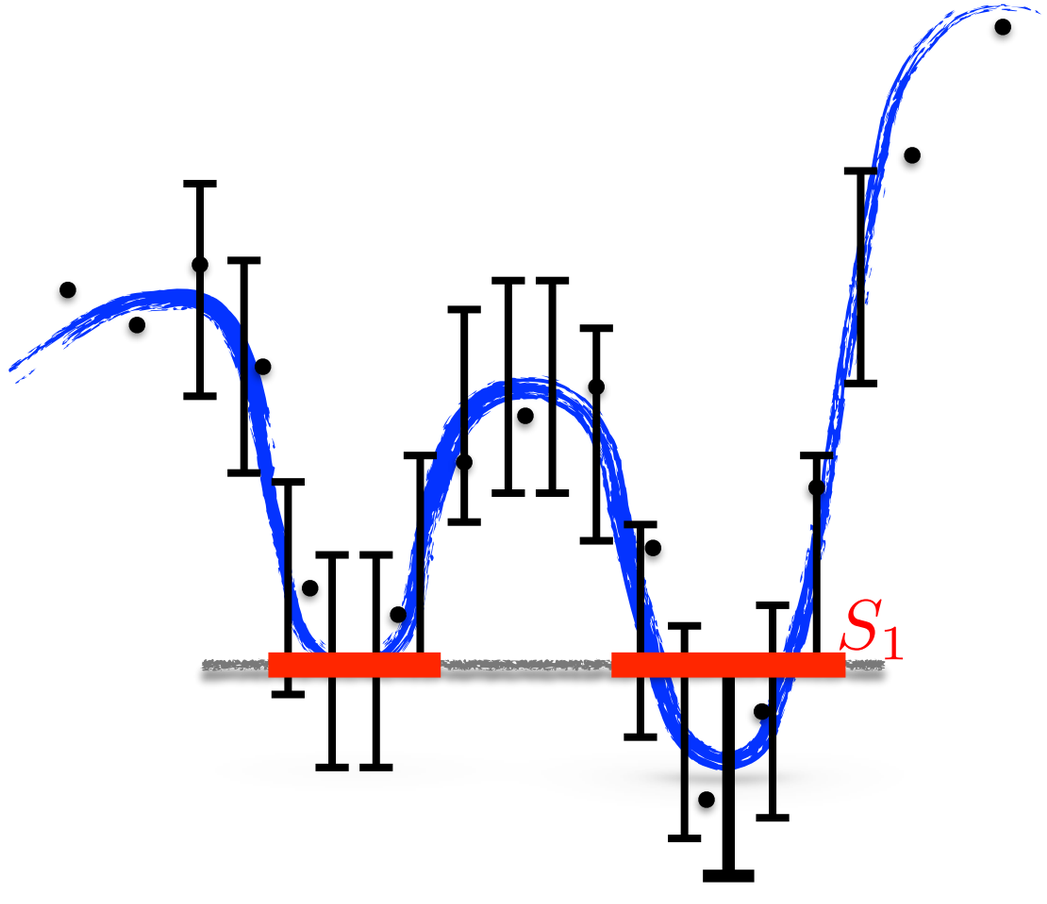
estimation error of f close to the reference f_0

PROOF SKETCH OF UPPER BOUND

- Our algorithm: "successive rejection"
  - *Step 1:* uniform sampling and build confidence intervals (CI)
  - *Step 2:* remove sub-optimal points
  - *Step 3:* uniform sample in the remaining points.

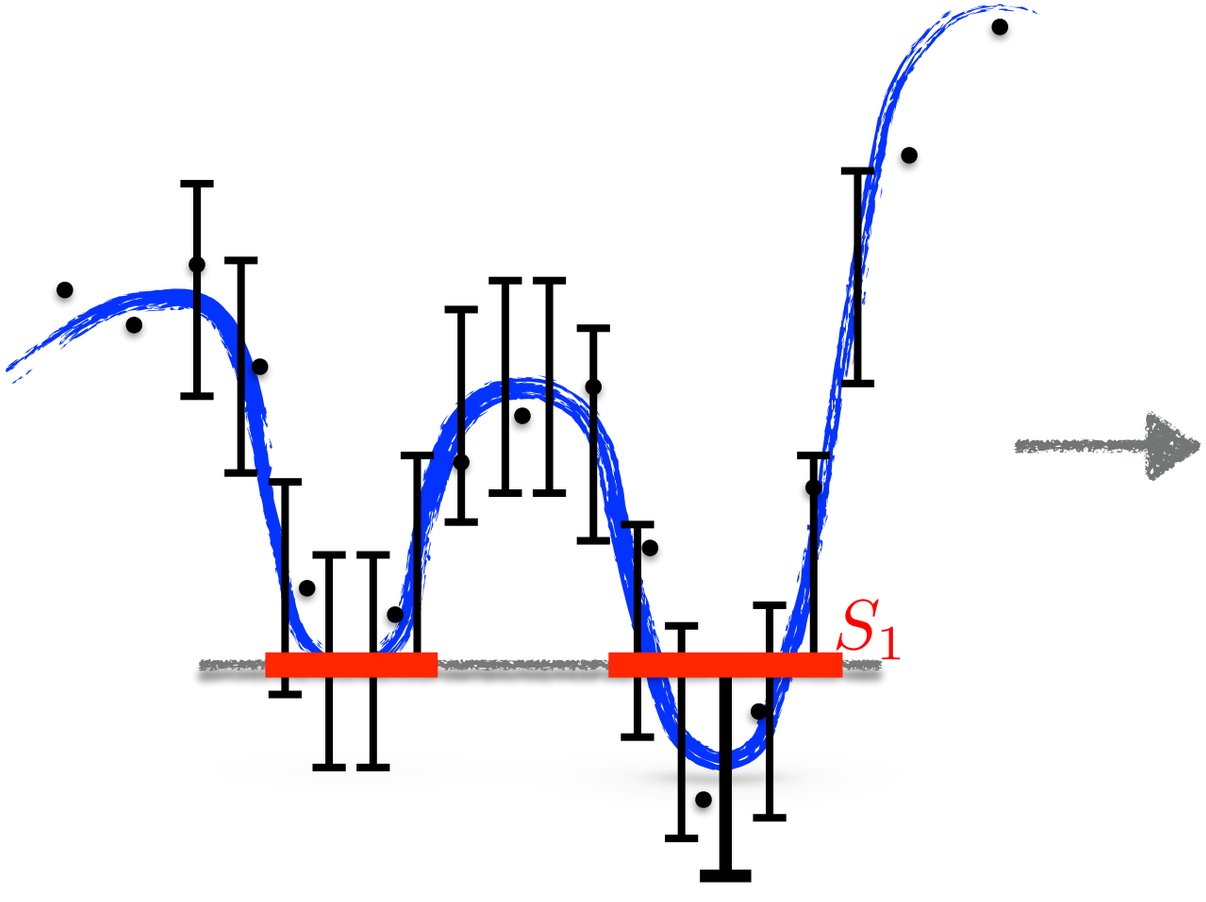
PROOF SKETCH OF UPPER BOUND

- Our algorithm: "successive rejection"
  - *Step 1:* uniform sampling and build confidence intervals (CI)
  - *Step 2:* remove sub-optimal points
  - *Step 3:* uniform sample in the remaining points.



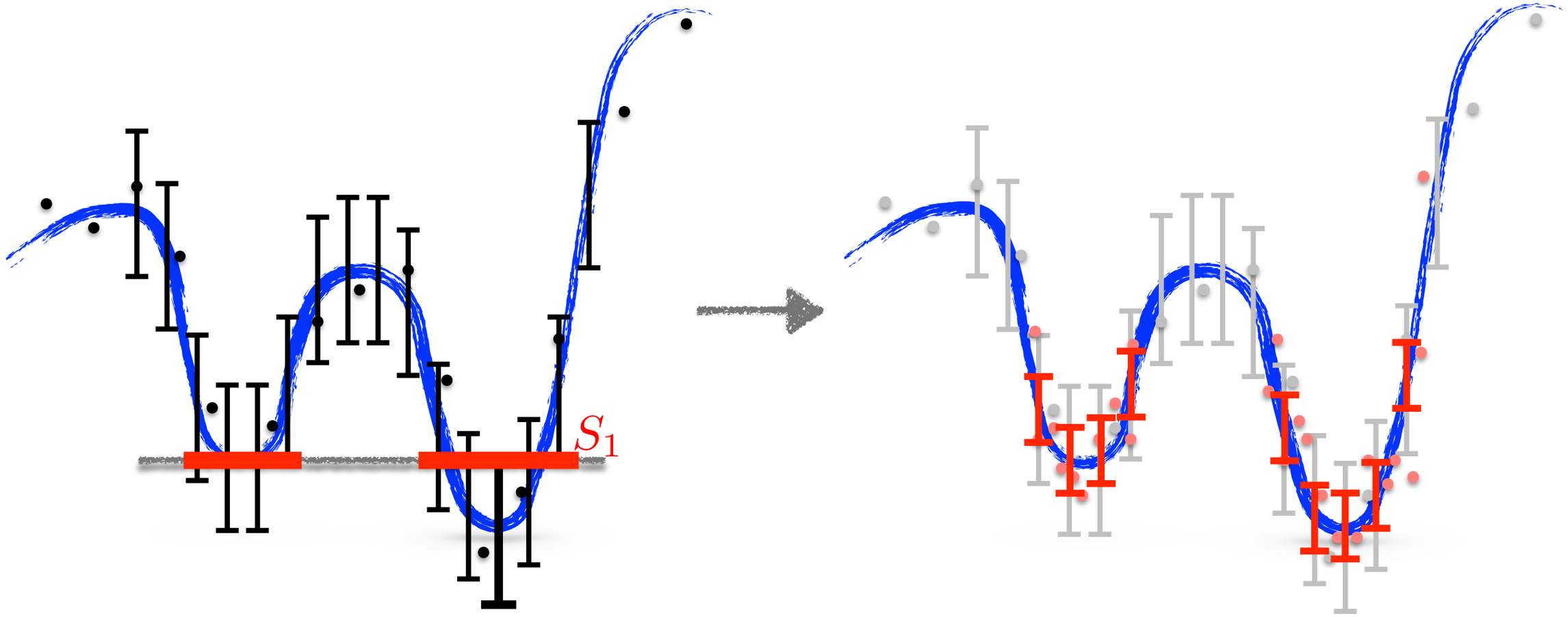
PROOF SKETCH OF UPPER BOUND

- Our algorithm: "successive rejection"
  - *Step 1:* uniform sampling and build confidence intervals (CI)
  - *Step 2:* remove sub-optimal points
  - *Step 3:* uniform sample in the remaining points.



PROOF SKETCH OF UPPER BOUND

- Our algorithm: "successive rejection"
  - *Step 1:* uniform sampling and build confidence intervals (CI)
  - *Step 2:* remove sub-optimal points
  - *Step 3:* uniform sample in the remaining points.



PROOF SKETCH OF UPPER BOUND

- Our algorithm: "successive rejection"
  - *Step 1:* uniform sampling and build confidence intervals (CI)
  - *Step 2:* remove sub-optimal points
  - *Step 3:* uniform sample in the remaining points.
- Key observation between iterations:

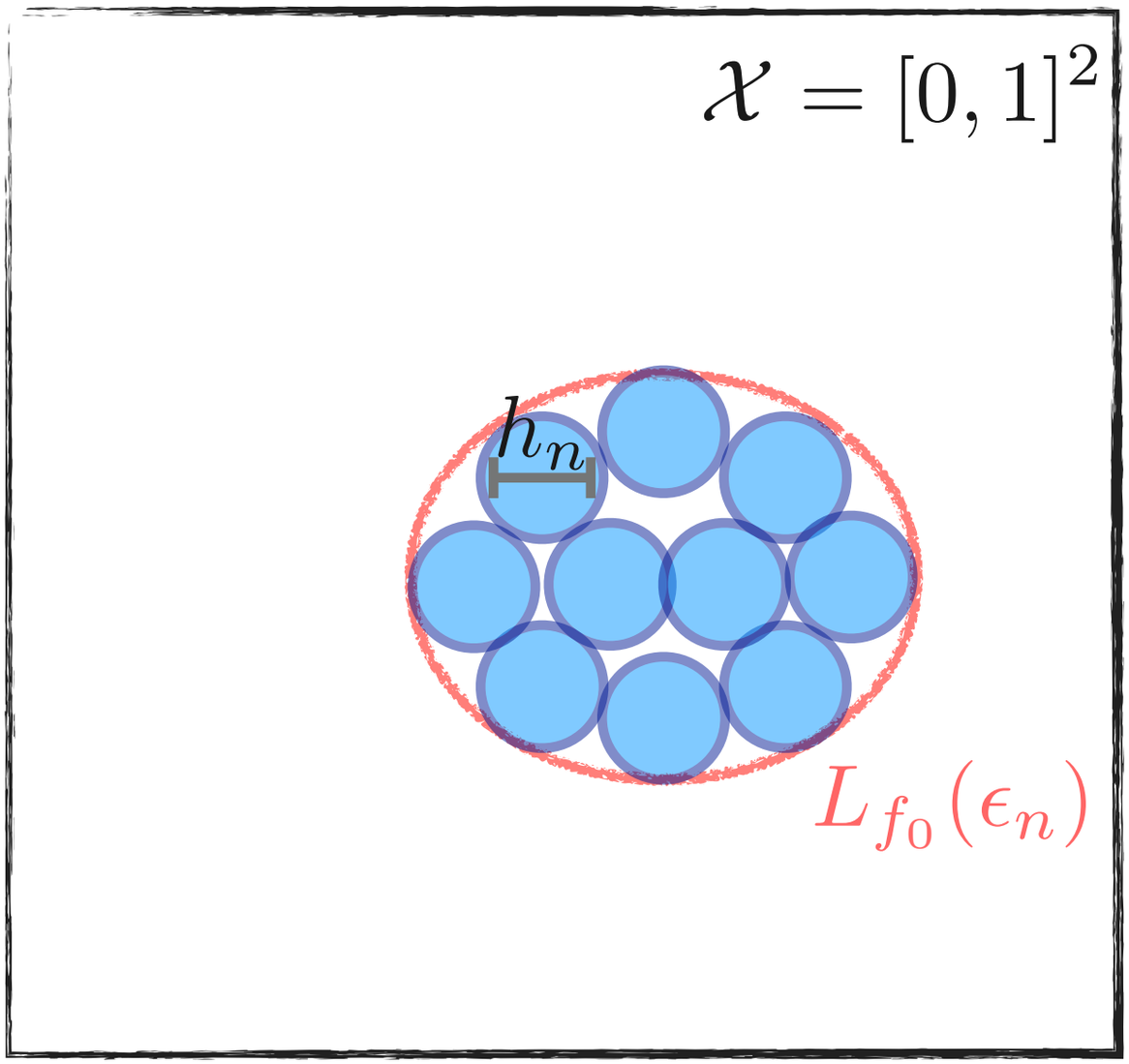
$$S_\tau \subseteq L_f(\varepsilon) \Longrightarrow S_{\tau+1} \subseteq L_f(\varepsilon/2)$$

$$\textit{Until} \quad \varepsilon \sim \varepsilon_n(f) \times \log^c n$$

- An $O(\log n)$ number of iterations suffice.

PROOF SKETCH OF LOWER BOUND

- Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



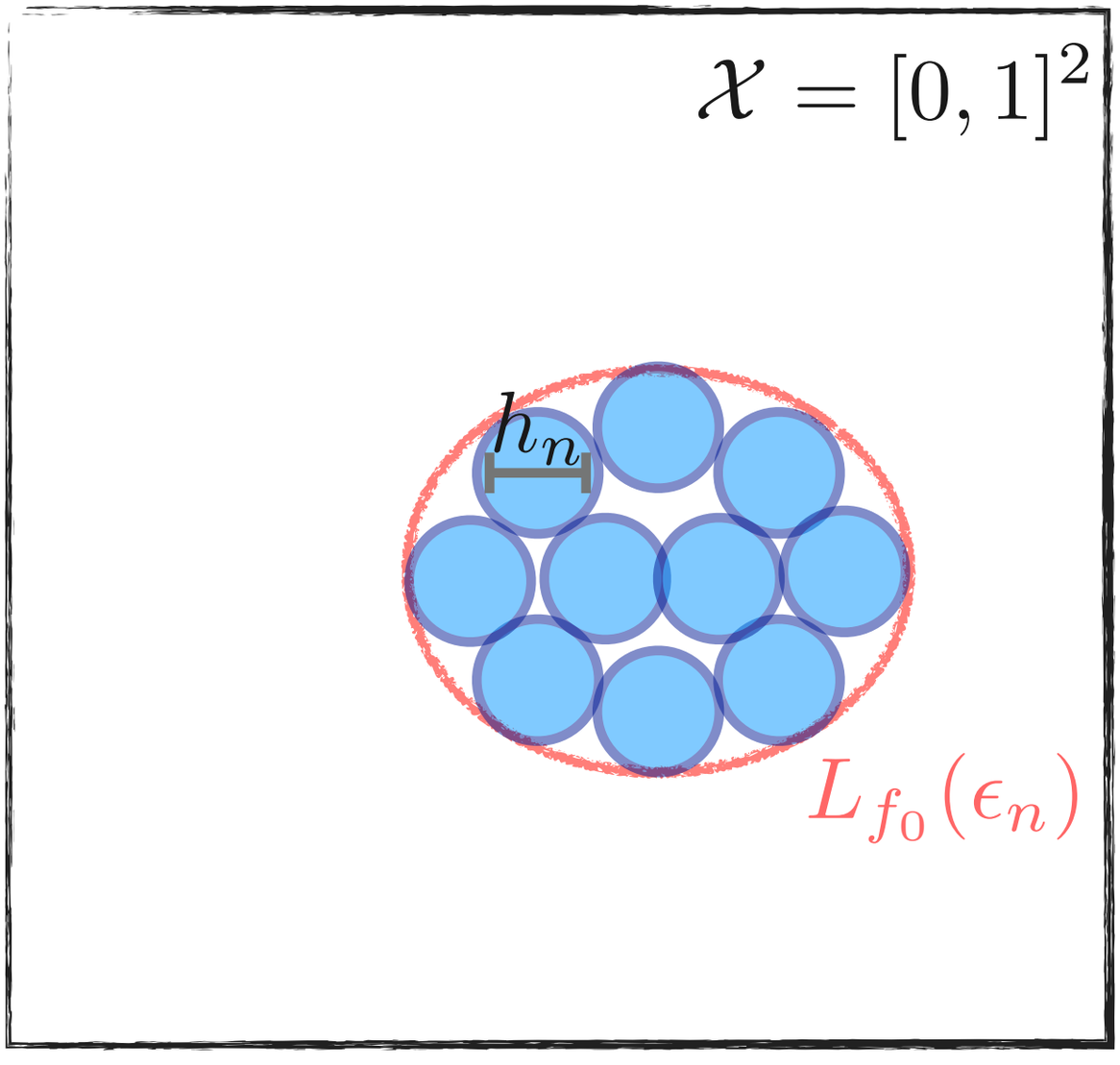
$h_n \asymp \epsilon_n^{1/\alpha}$

Discrepancy in ball: $2\epsilon_n$

Must identify the ball w.h.p.

PROOF SKETCH OF LOWER BOUND

➤ Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



Resembles

Bandit Pure Exploration

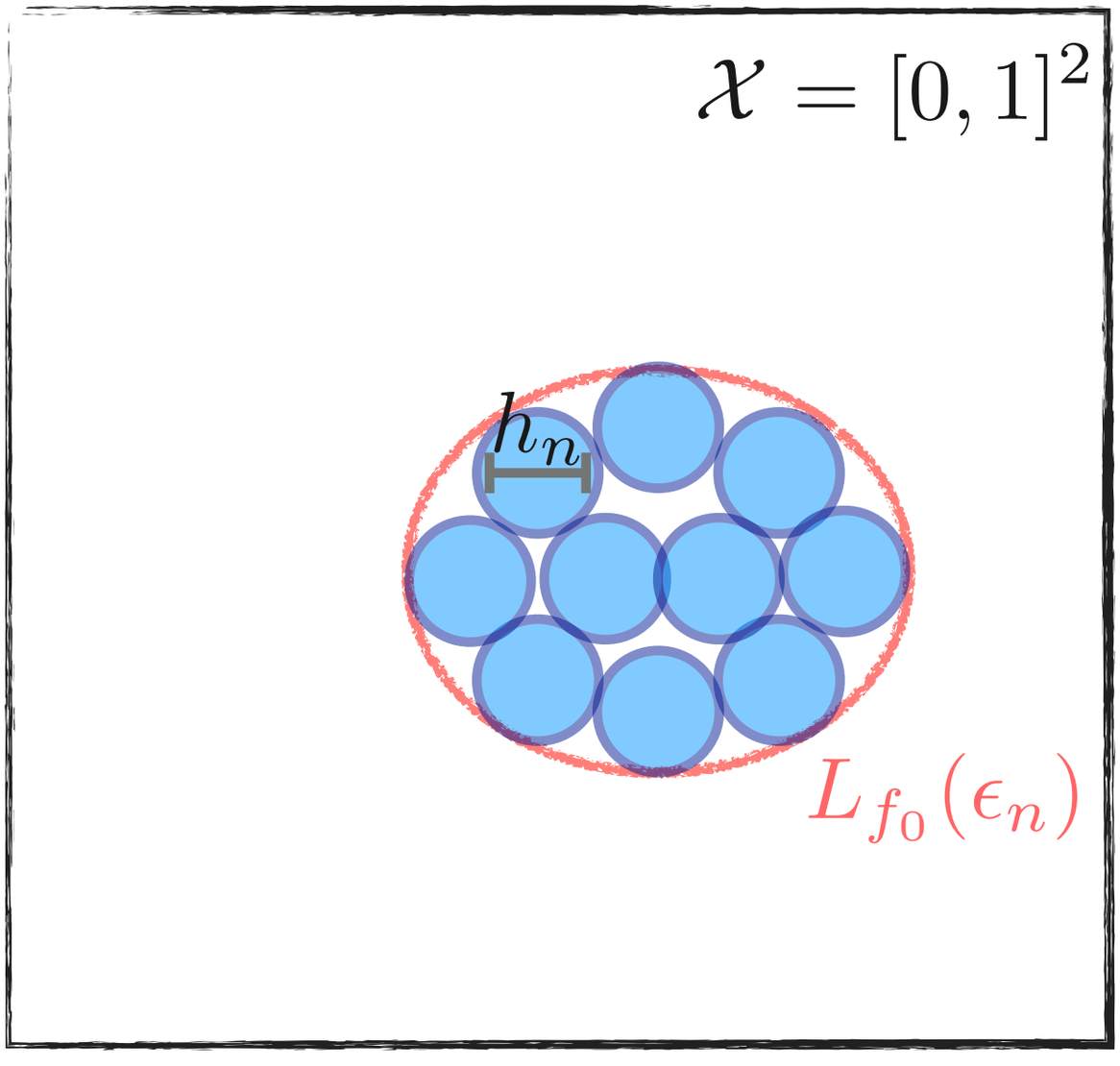
$\mu_1, \mu_2, \cdots, \mu_H \in \mathbb{R}$

$\mu_i = 2\epsilon_n; \mu_{-i} = 0$

Identify the non-zero arm

PROOF SKETCH OF LOWER BOUND

➤ Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



Resembles

Bandit Pure Exploration

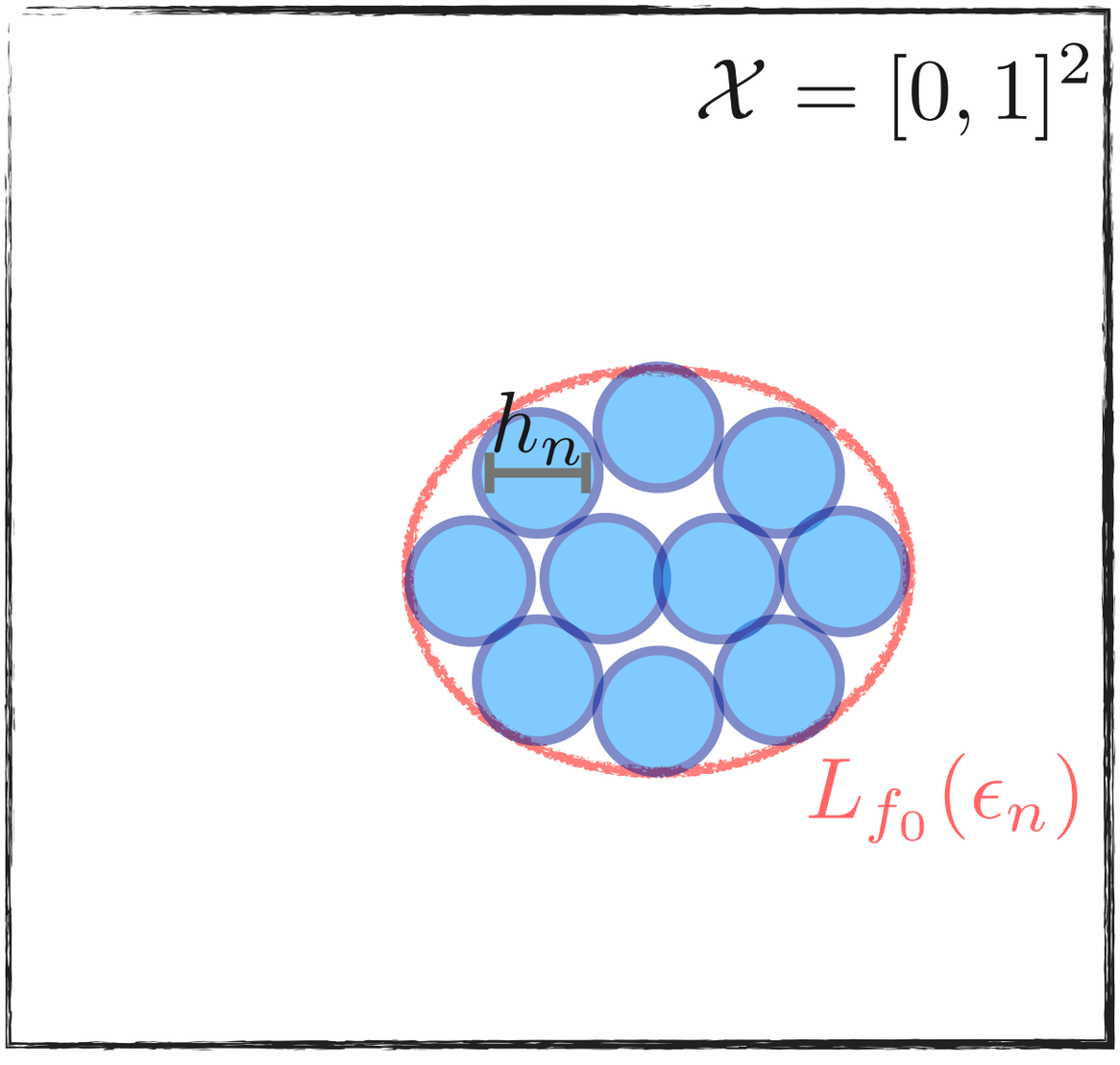
$$\mathrm{KL}(P_0 \| P_i) \lesssim n_i \cdot \epsilon_n^2$$

$$n_i \lesssim n/H$$

$$H \gtrsim \mu_{f_0}(\epsilon_n)/h_n^d$$

PROOF SKETCH OF LOWER BOUND

➤ Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



Resembles

Bandit Pure Exploration

$$\mathrm{KL}(P_0 \| P_i) \lesssim n_i \cdot \epsilon_n^2$$

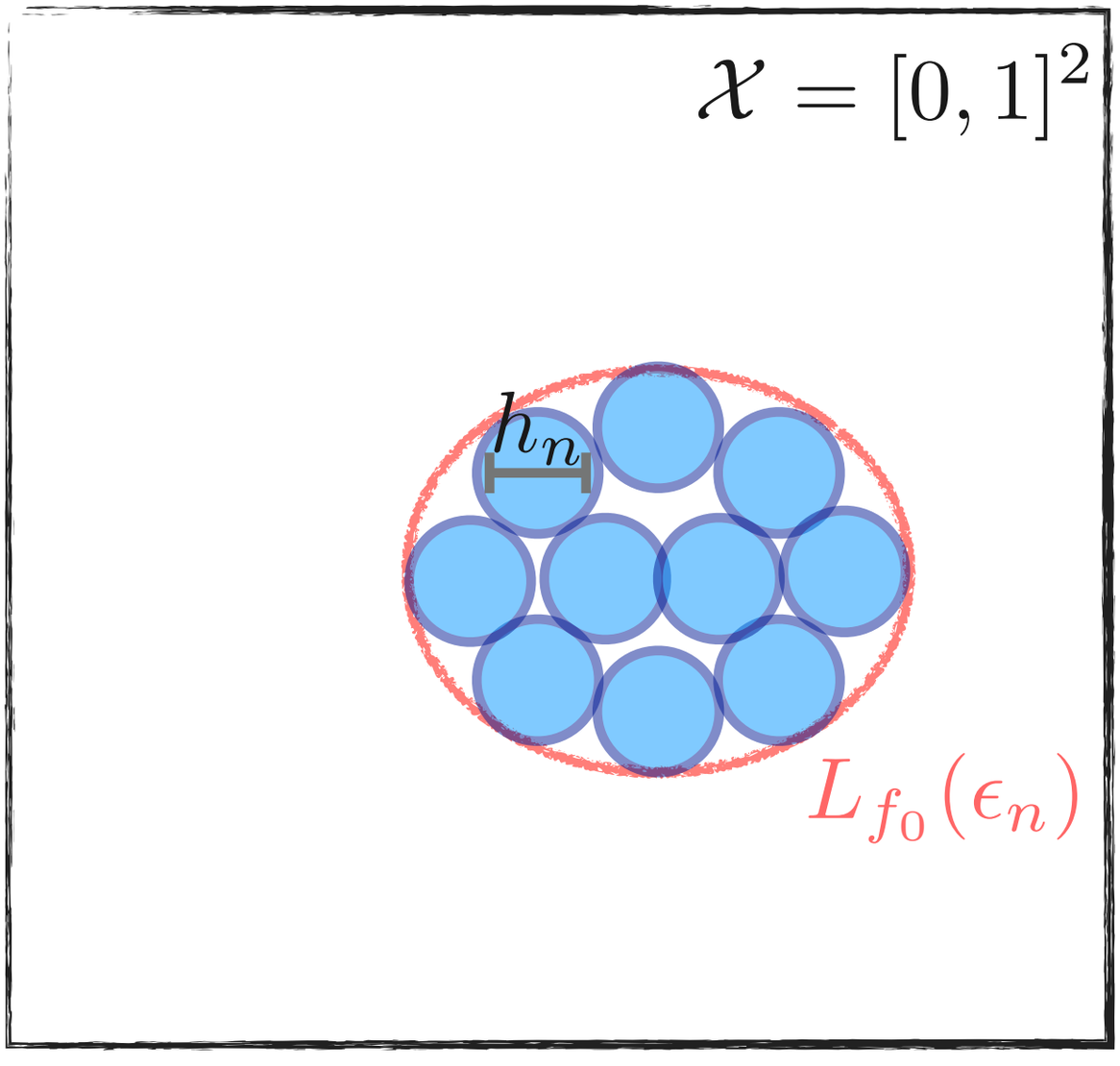
$$n_i \lesssim n/H$$

# of balls packed

$$H \gtrsim \mu_{f_0}(\epsilon_n)/h_n^d$$

PROOF SKETCH OF LOWER BOUND

➤ Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



Resembles

Bandit Pure Exploration

$$\mathrm{KL}(P_0 \| P_i) \lesssim n_i \cdot \epsilon_n^2$$

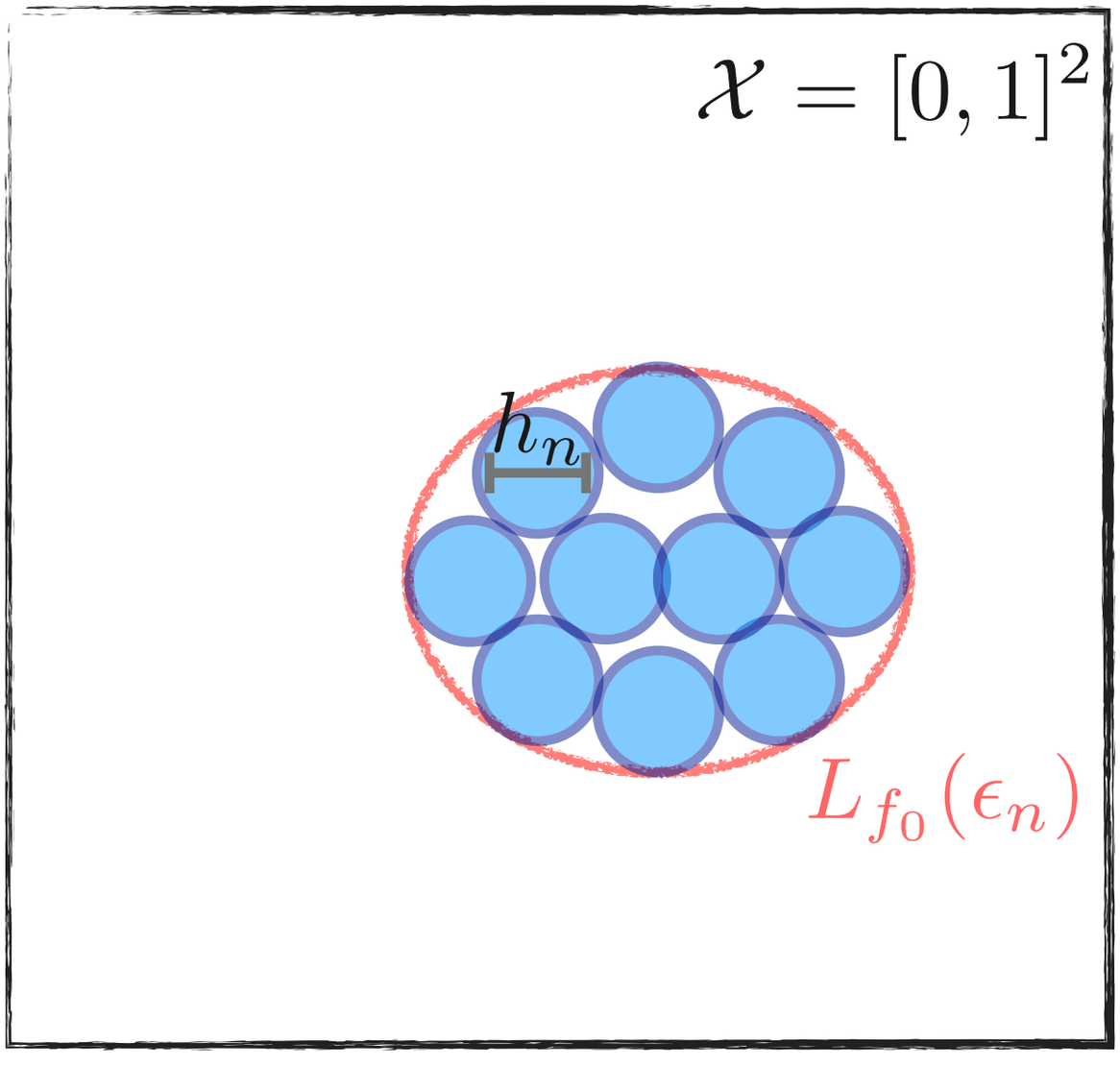
$$n_i \lesssim n/H$$

# of balls packed

$$H \gtrsim \mu_{f_0}(\epsilon_n)/h_n^d$$

PROOF SKETCH OF LOWER BOUND

➤ Step 1: constructing "packings" on $L_{f_0}(\epsilon_n)$



Resembles

Bandit Pure Exploration

$$\mathrm{KL}(P_0 \| P_i) \lesssim n_i \cdot \epsilon_n^2$$

$$n_i \lesssim n/H$$

# of balls packed

$$H \gtrsim \mu_{f_0}(\epsilon_n)/h_n^d$$

Regularity

TAKE-HOME MESSAGES

- (Noisy) zeroth-order optimization of smooth functions is in general difficult
  - ✳ As difficult as estimating the function in sup-norm.
- The optimal convergence rates exhibit significant gaps **locally** for different objective functions
  - ✳ Local minimax rate mostly dictated by level set growth;
  - ✳ The constant function is the hardest example;
  - ✳ Strongly convex functions do **not** exhibit curse of dim.
- A successive-rejection type algorithm is near-optimal.

FUTURE DIRECTIONS

- Are the regularity conditions absolutely necessary?
  - Can the level sets of f be irregular?
  - Can the volumes of level sets of f grow heterogeneously?
- Are there more computationally efficient algorithms?
  - Key challenge: avoiding creating sup-norm CIs explicitly.
- Log factors: are they removable? (conjecture: yes!)
  - Active queries methods for nonparametric estimation / bandit pure exploration do not have log terms.

QUESTIONS