Global Optimization with Active Learning

Thomas Moreau[1] Nicolas Vayatis[1]

[1]CMLA, ENS Cachan, CNRS, Université Paris Saclay, 94235 Cachan, France

Abstract

We introduce a random search algorithm for non-convex optimization. The algorithm is based on the estimation of a confidence set that contains the maximum of the function. The main steps of the estimation procedure are described and numerical experiments are presented to show how the algorithm works practically.

Keywords Optimization – Convolutational Sparse Coding – Time Series representation

1. Motivations and Problem Statement

Examples of High-dimensional Optimization Problems

- Optimization in Air Traffic Management
- Risk Minimization in Machine Learning
- Optimization of Energy Networks

Derivative-free Optimization [2]

Let $f^\star : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subset \mathbb{R}^d$ is a closed set

- Estimating $\arg\max_{x\in\mathcal{X}} f^\star(x) = \{x \in \mathcal{X} : \forall x' \in \mathcal{X}, f^\star(x') \le f^\star(x)\}$
- Sequential queries $(x_1, f^\star(x_1)), (x_2, f^\star(x_2)), \dots$
- Choosing x_{n+1} using the previous observations $\mathcal{D}_n = \{(x_i, f^\star(x_i))\}_{i\le n}$

Main Objectives

- Design a non-greedy algorithm for non-convex optimization
- Keep the computational cost low for large dimensions (d $\gg$ 10)
- Confirm the performance empirically
- Provide theoretical guarantees

Hypothesis and Assumptions

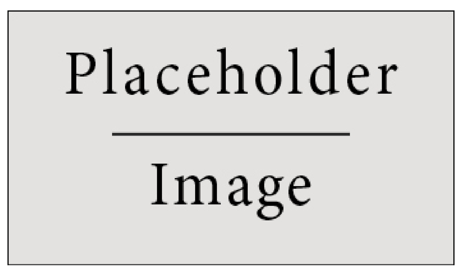
- The unkown function $f^\star$ is continuous
- The unkown function has no flat parts
- The unkown function has reasonable variations

2. Controlling the Smoothness

Bounding the variations

To enforce near-by locations to have close associated values, we assume that $f^\star$ belongs to a carefully chosen set of functions $\mathcal{F}$. We consider the following sets of functions:

- Lipschitzian functions: $|f(x) - f(x')| \le k \, \|x - x'\|_2$
- Gaussian RKHS balls: $|f(x) - f(x')| \le M \left(1 - \exp\left(-\frac{\|x-x'\|_2^2}{2\sigma^2}\right)\right)^{1/2}$



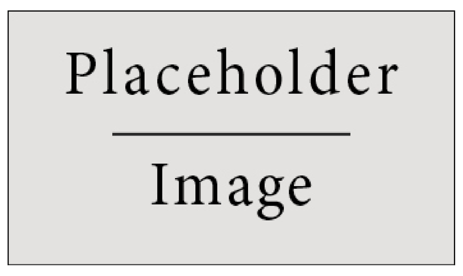


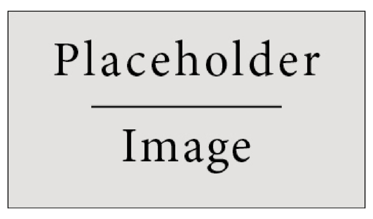
Figure: Upper and lower bounds for lipschizian functions (left) and gaussian RKHS (right)

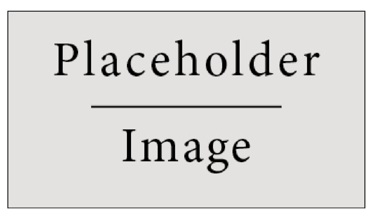
3. Concepts of Active Learning

Active learning for binary classification [1]

Let $h^\star \in \mathcal{H}$ and $\mathcal{H}_n = \{h \in \mathcal{H} : \sum_{i=1}^n 1\{h(X_i) \ne h^\star(X_i)\} = 0\}$

- $DIS(\mathcal{H}_n) = \{x \in \mathcal{X} : \exists (h, h') \in \mathcal{H}_n^2,\ h(x) \ne h'(x)\}$
- $SUP(\mathcal{H}_n) = \{x \in \mathcal{X} : \forall h \in \mathcal{H}_n,\ h(x) = 1\}$
- $INF(\mathcal{H}_n) = \{x \in \mathcal{X} : \forall h \in \mathcal{H}_n,\ h(x) = 0\}$





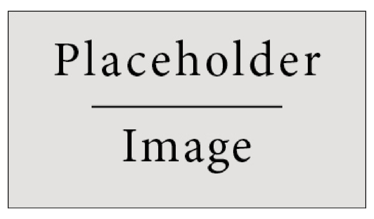


Figure: Computation of the disagreement area for linear classifiers

4. Estimating the Maximum with Active Learning

Set of candidates

Given a set of mpoints $\{(X_i, f^\star(X_i))\}_{i\le n}$ and a functional space $\mathcal{F}$, we define the set of functions that are still consistent with the datas:

$$\mathcal{F}_n = \{f \in \mathcal{F} : \forall i \in \{1 \dots n\},\ f(X_i) = f^\star(X_i)\}.$$

Query procedure

To keep the consistency of the optimization procedure, we propose a query policy that consists in selecting points uniformely from:

$$\textsc{ArgMax}(\mathcal{F}_n) = \{x \in \mathcal{X} :\ \exists f \in \mathcal{F}_n \text{ s.t. } x \in \arg\max_{x\in\mathcal{X}} f(x)\}.$$

Lipschitzian case

If $\mathcal{F}$ =Lip(k), we have that

$x \in \textsc{ArgMax}(\mathcal{F}_n)$ i.f.f. $\min_{i\le n} f^\star(X_i) + k\,\|x - X_i\|_2 \ge \max_{i\le n} f^\star(X_i)$.

5. Selective Algorithm

Input: $\mathcal{F}$, $\mathcal{X}$, n

Init: $t, m \leftarrow 0$, $\mathcal{F}_0 \leftarrow \mathcal{F}$, $X_1, X_2, \dots \overset{\text{iid}}{\sim} \mathcal{U}(\mathcal{X})$

While $t < n$

  $m \leftarrow m + 1$

  If $X_m \in \textsc{ArgMax}(\mathcal{F}_t)$

    Request $f^\star(X_m)$, $t \leftarrow t + 1$

    $\mathcal{F}_t \leftarrow \{f \in \mathcal{F}_{t-1} : f(X_m) = f^\star(X_m)\}$

Output: Any $x \in \textsc{ArgMax}(\mathcal{F}_n)$

6. Illustrations

Figure: Selective Algortihm on four synthetic functions

7. Open Questions and Discussion

- Theoretical performance of the algorithm
- Comparison with natural competitors
- Robustness to noise

References

[1] J. Doe and J. Smith. A random citation? *arXiv preprint arXiv:xxx.xxxx*, (1):1–9, 2015.

[2] J. Doe, J. Smith, and J. Average. Yet another convolutional neural net. In *Advances in Neural Information Processing Systems (NIPS)*, pages 577—585, 2025.