Stochastic Simultaneous Optimistic Optimization

Michal.Valko@inria.fr, A.Carpentier@statslab.cam.ac.uk, and Rémi.Munos@inria.fr

Setting

StoSOO is a global function maximizer:

- **Goal:** Maximize $f : \mathcal{X} \to I\!R$ given a budget of n evaluations.
- **Challenges:** f is ***stochastic*** and has ***unknown smoothness***
- **Protocol:** At round t, select state x_t, observe r_t such that

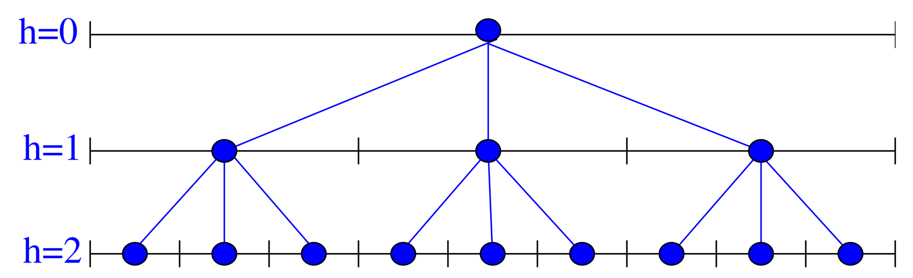
$$\mathbb{E}[r_t|x_t] = f(x_t).$$

After n rounds, return a state $x(n)$.

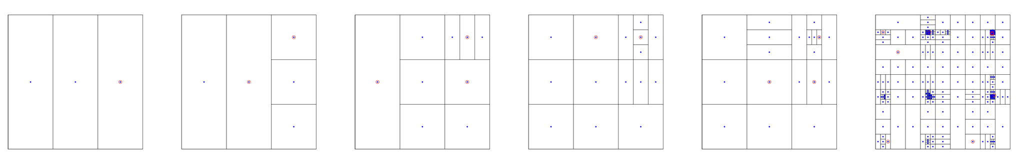
- **Loss:** $R_n = \sup_{x\in\mathcal{X}} f(x) - f(x(n))$

StoSOO operates on a given **hierarchical partitioning** of $\mathcal{X}$:

- For any h, $\mathcal{X}$ is partitioned in K^h cells $(X_{h,i})_{0\le i\le K^h-1}$.
- K-ary tree $\mathcal{T}_\infty$ where depth $h = 0$ is the whole $\mathcal{X}$.



- StoSOO adaptively creates finer and finer partitions of $\mathcal{X}$.



- $x_{h,i} \in X_{h,i}$ is a specific state per cell where f is evaluated

Comparison

	deterministic	stochastic
known smoothness	DOO	Zooming or HOO
unknown smoothness	DIRECT or SOO	StoSOO

Hierarchical optimistic optimization algorithms

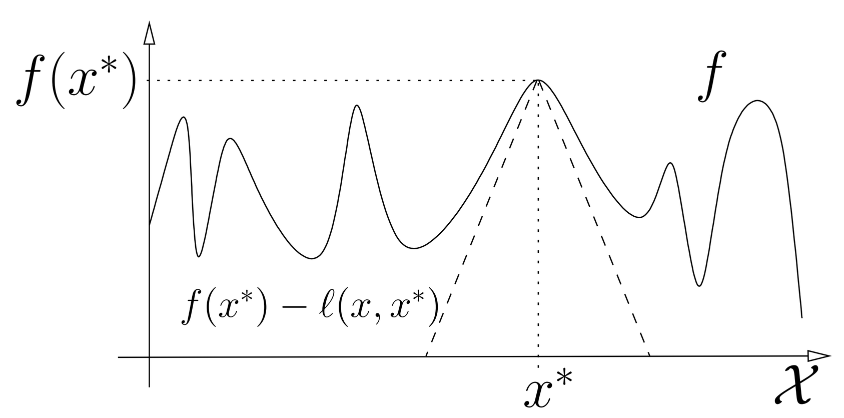
Assumptions

There exists a semi-metric ℓ on $\mathcal{X}$ (triangle inequality not required):

A1 Local smoothness of f: For all $x \in \mathcal{X}$:

$$f(x^*) - f(x) \le \ell(x, x^*).$$

"f does not decrease too fast around x^*"



A2 Bounded diameters and well-shaped cells: There exists a decreasing sequence $w(h) > 0$, such that for any depth $h \ge 0$ and for any cell $\mathcal{X}_{h,i}$ of depth h, we have $\sup_{x\in X_{h,i}} \ell(x_{h,i}, x) \le w(h)$. Moreover, there exists $\nu > 0$ such that for any depth $h \ge 0$, any cell $\mathcal{X}_{h,i}$ contains a ℓ-ball of radius $\nu w(h)$ centered in $x_{h,i}$.

Measure of complexity

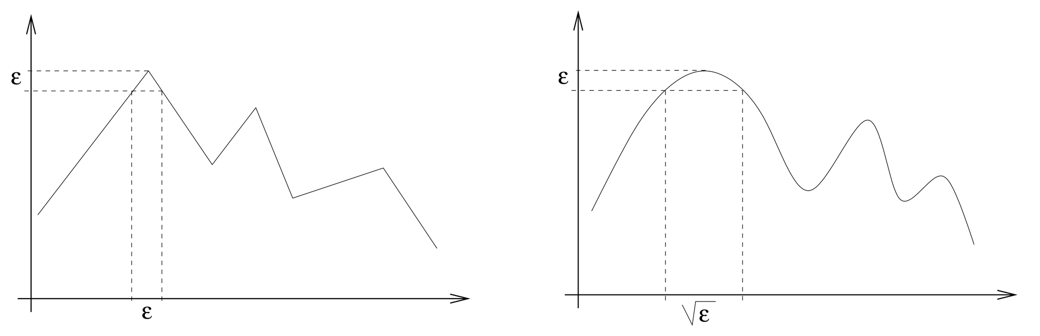
For any $\varepsilon > 0$, write the set of ε-optimal states:

$$\mathcal{X}_\varepsilon \stackrel{\text{def}}{=} \{x \in \mathcal{X}, f(x) \ge f^* - \epsilon\}$$

Definition 1 (near-optimality dimension). *Smallest constant d such that there exists $C > 0$, for all $\varepsilon > 0$, the packing number of $\mathcal{X}_\varepsilon$ with ℓ-balls of radius $\nu\epsilon$ is less than $C\varepsilon^{-d}$.*

Illustration:

$f(x^*) - f(x) = \Theta(||x^* - x||)$ $\quad$ $f(x^*) - f(x) = \Theta(||x^* - x||^2)$



$\ell(x, y) = ||x - y|| \implies d = 0$ $\quad$ $\ell(x, y) = ||x-y|| \implies d = D/2$
$\ell(x, y) = ||x - y||^2 \implies d = 0$

StoSOO Algorithm

Parameters: number of function evaluations n, maximum number of evaluations per node $k > 0$, maximum depth $h_{\max}$, and $\delta > 0$.
Initialization:
  $\mathcal{T} \leftarrow \{\circ[0, 0]\}$ {root node}
  $t \leftarrow 0$ {number of evaluations}
  $m \leftarrow 0$ {number of leaf expansions}
while $t \le n$ **do**
  $b_{\max} \leftarrow -\infty$
  for $h = 0$ to $\min(\text{depth}(\mathcal{T}), h_{\max})$ **do**
    if $t \le n$ **then**
      For each leaf $\circ[h, j] \in \mathcal{L}$, compute its b-value:
      $b_{h,j}(t) = \hat{\mu}_{h,j}(t) + \sqrt{\log(nk/\delta)/(2T_{h,j}(t))}$
      Among leaves $\circ[h, j] \in \mathcal{L}_t$ at depth h, select

$$\circ[h, i] \in \mathop{\arg\max}_{\circ[h,j]\in\mathcal{L}} b_{h,j}(t)$$

      if $b_{h,i}(t) \ge b_{\max}$ **then**
        if $T_{h,i}(t) < k$ **then**
          Evaluate (sample) state $x_t = x_{h,i}$.
          Collect reward r_t (s.t. $\mathbb{E}[r_t|x_t] = f(x_t)$).
          $t \leftarrow t + 1$
        else {i.e. $T_{h,i}(t) \ge k$, expand this node}
          Add the K children of $\circ[h, i]$ to $\mathcal{T}$
          $b_{\max} \leftarrow b_{h,i}(t)$
        end if
      end if
    end if
  end for
end while
Output: The representative point with the highest $\hat{\mu}_{h,j}(n)$ among the deepest expanded nodes:

$$x(n) = \mathop{\arg\max}_{x_{h,j}} \hat{\mu}_{h,j}(n) \text{ s.t. } h = \text{depth}(\mathcal{T} \setminus \mathcal{L}).$$

How it works?

- StoSOO iteratively traverses and builds a tree over $\mathcal{X}$
- at each traversal it selects several nodes **simultaneously**
- the selection is **optimistic**, based on confidence bounds
- selected nodes are either **sampled** or **expanded**
- **sample** the node k times for a confident estimate of $f(x_{h,i})$
- returns the deepest **expanded** node

Analysis

Main result:

Theorem 1. *Let d be the $\nu/3$-near-optimality dimension and C be the corresponding constant. If the assumptions hold, then the loss of StoSOO run with parameters k, $h_{\max}$, and $\delta > 0$, after n iterations is bounded, with probability $1 - \delta$, as:*

$$R_n \le 2\varepsilon + w\left(\min\left(h(n) - 1, h_\varepsilon, h_{\max}\right)\right)$$

where $\varepsilon = \sqrt{\log(nk/\delta)/(2k)}$ *and $h(n)$ is the smallest $h \in \mathbb{N}$, such that:*

$$C(k + 1)h_{\max} \sum_{l=0}^{h} (w(l) + 2\varepsilon)^{-d} \ge n,$$

and h_ε is defined as:

$$h_\varepsilon = \arg\min\{h \in \mathbb{N} : w(h + 1) < \varepsilon\}.$$

Exponential diameters and $d = 0$**:**

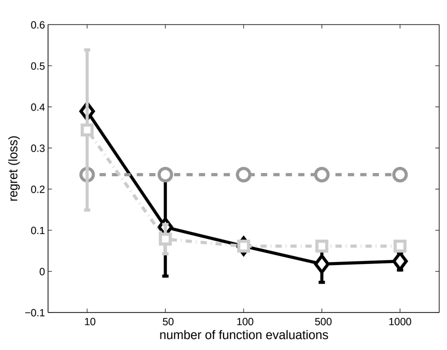
Corollary 1. *Assume that the diameters of the cells decrease exponentially fast, i.e., $w(h) = c\gamma^h$ for some $c > 0$ and $\gamma < 1$. Assume that the $\nu/3$-near-optimality dimension is $d = 0$ and let C be the corresponding constant. Then the expected loss of StoSOO run with parameters k, $h_{\max} = \sqrt{n/k}$, and $\delta > 0$, is bounded as:*

$$\mathbb{E}[R_n] \le (2 + 1/\gamma)\varepsilon + c\gamma^{\sqrt{n/k}\min\{0.5/C,1\}-2} + 2\delta.$$

Corollary 2. *For the choice $k = n/\log^3(n)$ and $\delta = 1/\sqrt{n}$, we have:*

$$\mathbb{E}[R_n] = O\left(\frac{\log^2(n)}{\sqrt{n}}\right).$$

This result shows that, surprisingly, StoSOO can achieve the same rate $\tilde{O}(n^{-1/2})$, up to a logarithmic factor, as the HOO or Stochastic DOO algorithms run with the best possible metric, although StoSOO does not require the knowledge of it. StoSOO (diamonds) vs. Stochastic DOO with ℓ_1 (circles) and ℓ_2 (squares) on f_1.

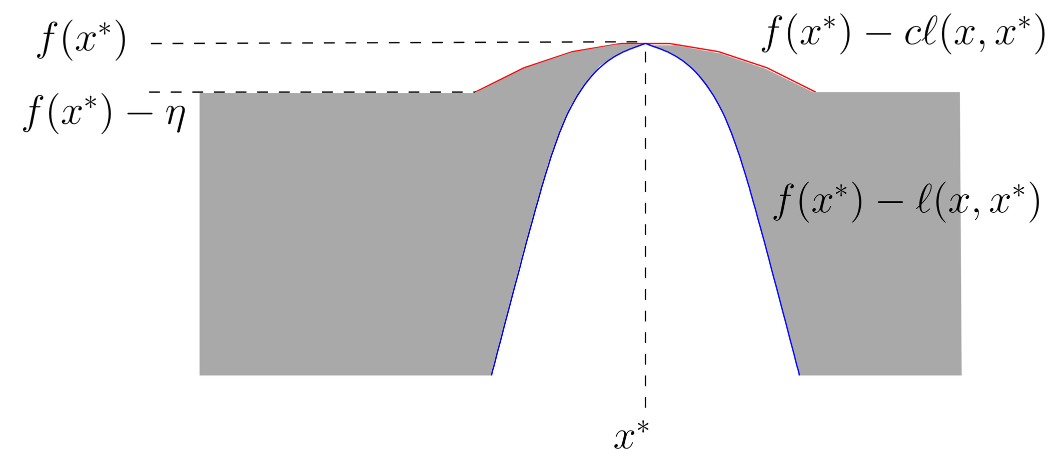


The important case $d = 0$

Example 1: Functions f defined on $[0, 1]^D$ that are locally equivalent to a polynomial of degree α around their maximum, i.e., $f(x) - f(x^*) = \Theta(\|x - x^*\|^\alpha)$ for some $\alpha > 0$, where $\|\cdot\|$ is any norm. The choice of semi-metric $\ell(x, y) = \|x - y\|^\alpha$ implies that the near-optimality dimension $d = 0$. This covers already a large class of functions.

Example 2: More generally, we consider a finite dimensional and bounded space $\mathcal{X}$, (e.g., Euclidean space $[0, 1]^D$) with a finite doubling constant. Let a function in such space have upper- and lower envelope around x^* of the same order, i.e., there exists constants $c \in (0, 1)$, and $\eta > 0$, such that for all $x \in \mathcal{X}$:

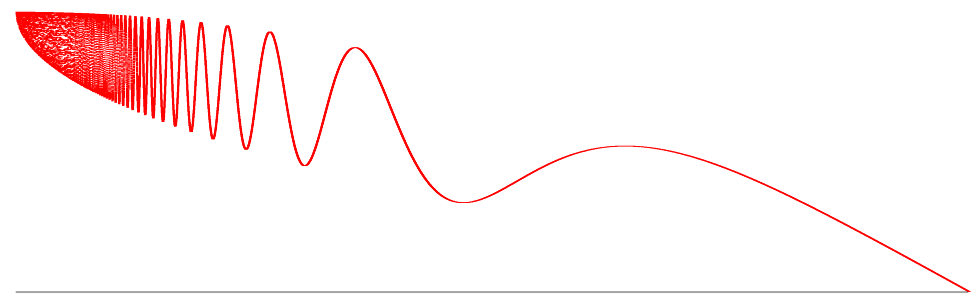
$$\min(\eta, c\ell(x, x^*)) \le f(x^*) - f(x) \le \ell(x, x^*). \quad (1)$$



Any function satisfying (1) lies in the gray area and possesses a lower- and upper-envelopes that are of same order around x^*.

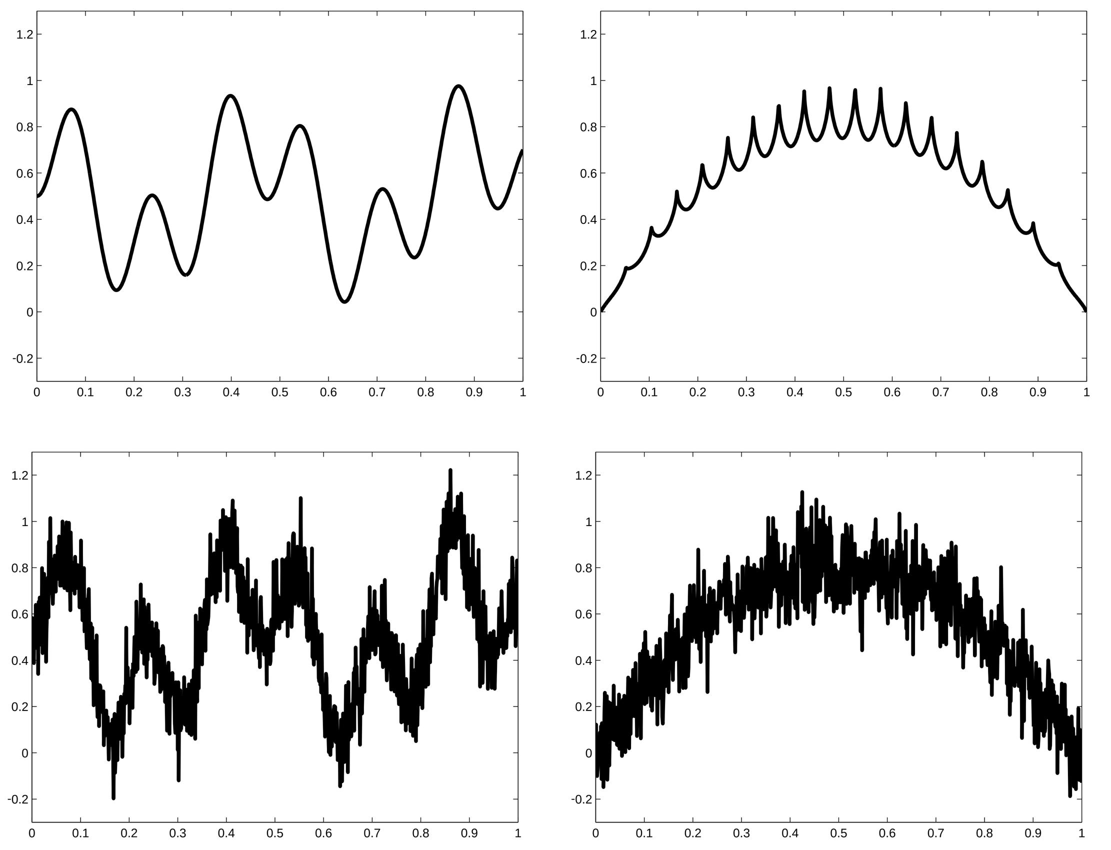
Example of a function with different order in the upper and lower envelopes, when $\ell(x, y) = |x - y|^\alpha$:

$$f(x) = 1 - \sqrt{x} + (-x^2 + \sqrt{x}) \cdot (\sin(1/x^2) + 1)/2$$

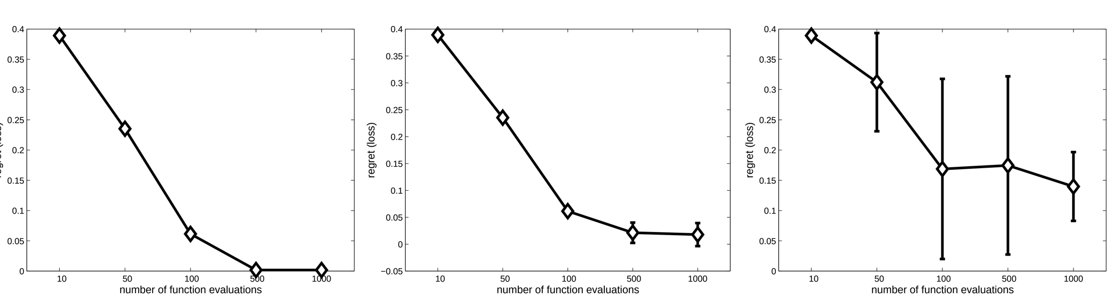


The lower-envelope behaves like a square root whereas the upper one is quadratic. The maximum number of ℓ-balls with radius ε that can pack $\mathcal{X}_\varepsilon$ (i.e., Euclidean balls with radius $\varepsilon^{1/\alpha}$) is at most of order $\varepsilon^{1/2}/\varepsilon^{1/\alpha} \le \varepsilon^{-3/2}$, since $\alpha \le 1/2$ in order to satisfy the assumption on f. We deduce that there is no semi-metric of the form $|x - y|^\alpha$ for which $d < 3/2$.

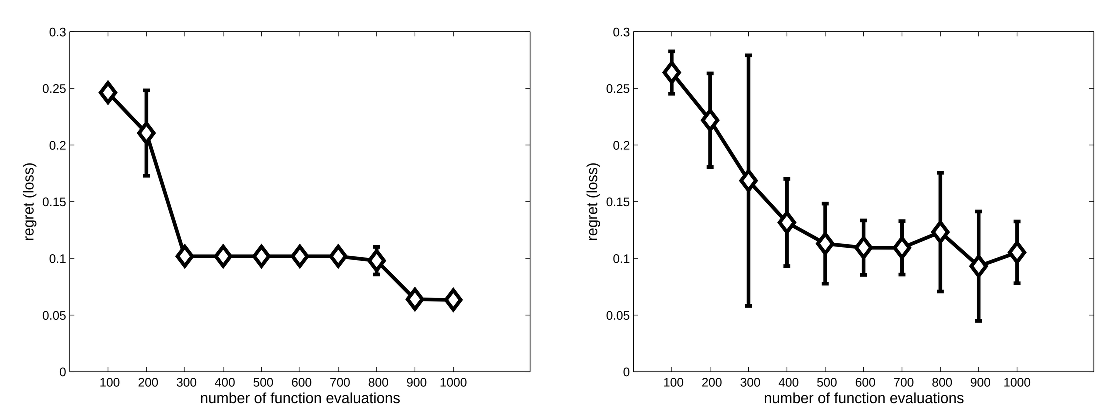
Experiments



Left: *Two-sine product* function $f_1(x) = \frac{1}{2}(\sin(13x) \cdot \sin(27x)) + 0.5$.
Right: *Garland* function: $f_2(x) = 4x(1-x)\cdot(\frac{3}{4} + \frac{1}{4}(1 - \sqrt{|\sin(60x)|}))$.



StoSOO's on f_1. **Left**: Noised with $\mathcal{N}_T(0, 0.01)$. **Middle**: Noised with $\mathcal{N}_T(0, 0.1)$. **Right**: Noised with $\mathcal{N}_T(0, 1)$.



StoSOO's performance for the garland function. **Left** noised with $\mathcal{N}_T(0, 0.01)$. **Right**: Noised with $\mathcal{N}_T(0, 0.1)$.

Code at: https://sequel.lille.inria.fr/Software/StoSOO