Bandit View on Continuous Stochastic Optimization

Sébastien Bubeck\textsuperscript{1}

\textit{joint work with} Rémi Munos\textsuperscript{1} & Gilles Stoltz\textsuperscript{2} & Csaba Szepesvari\textsuperscript{3}

\textsuperscript{1} INRIA Lille, SequeL team
\textsuperscript{2} CNRS/ENS/HEC
\textsuperscript{3} University of Alberta
\( \mathcal{X} \)-armed bandit game

**Parameters available to the forecaster:** the number of rounds \( n \) and the set of arms \( \mathcal{X} \).

**Parameters unknown to the forecaster:** mean-payoff function \( f : \mathcal{X} \to [0, 1] \), reward distributions (over \([0, 1]) \) \( M(x) \) such that \( f(x) \) is the expectation of \( M(x) \).

For each round \( t = 1, 2, \ldots, n \):

1. The player chooses an arm \( X_t \in \mathcal{X} \).
2. The environment draws the reward \( Y_t \) from \( M(X_t) \) (and independently from the past given \( X_t \)).

**Goal:** Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

\[
R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).
\]
\(\mathcal{X}\)-armed bandit game

**Parameters available to the forecaster:** the number of rounds \(n\) and the set of arms \(\mathcal{X}\).

**Parameters unknown to the forecaster:** mean-payoff function \(f : \mathcal{X} \to [0, 1]\), reward distributions (over \([0, 1]\)) \(M(x)\) such that \(f(x)\) is the expectation of \(M(x)\).

For each round \(t = 1, 2, \ldots, n\):

1. The player chooses an arm \(X_t \in \mathcal{X}\).
2. The environment draws the reward \(Y_t\) from \(M(X_t)\) (and independently from the past given \(X_t\)).

**Goal:** Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

\[
R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).
\]
\( \mathcal{X} \)-armed bandit game

**Parameters available to the forecaster:** the number of rounds \( n \) and the set of arms \( \mathcal{X} \).

**Parameters unknown to the forecaster:** mean-payoff function \( f : \mathcal{X} \to [0, 1] \), reward distributions (over \([0, 1]\)) \( M(x) \) such that \( f(x) \) is the expectation of \( M(x) \).

For each round \( t = 1, 2, \ldots, n \):

1. The player chooses an arm \( X_t \in \mathcal{X} \).
2. The environment draws the reward \( Y_t \) from \( M(X_t) \) (and independently from the past given \( X_t \)).

**Goal:** Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

\[
R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).
\]
\(\mathcal{X}\)-armed bandit game

**Parameters available to the forecaster:** the number of rounds \(n\) and the set of arms \(\mathcal{X}\).

**Parameters unknown to the forecaster:** mean-payoff function \(f : \mathcal{X} \rightarrow [0, 1]\), reward distributions (over \([0, 1]\)) \(M(x)\) such that \(f(x)\) is the expectation of \(M(x)\).

For each round \(t = 1, 2, \ldots, n\);  

1. The player chooses an arm \(X_t \in \mathcal{X}\).
2. The environment draws the reward \(Y_t\) from \(M(X_t)\) (and independently from the past given \(X_t\)).

**Goal:** Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

\[
R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).
\]
\(\mathcal{X}\)-armed bandit game

Parameters available to the forecaster: the number of rounds \(n\) and the set of arms \(\mathcal{X}\).

Parameters unknown to the forecaster: mean-payoff function \(f : \mathcal{X} \to [0, 1]\), reward distributions (over \([0, 1]\)) \(M(x)\) such that \(f(x)\) is the expectation of \(M(x)\).

For each round \(t = 1, 2, \ldots, n\):

1. The player chooses an arm \(X_t \in \mathcal{X}\).
2. The environment draws the reward \(Y_t\) from \(M(X_t)\) (and independently from the past given \(X_t\)).

Goal: Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

\[
R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).
\]
Parameters available to the forecaster: the number of rounds $n$ and the set of arms $\mathcal{X}$.

Parameters unknown to the forecaster: mean-payoff function $f : \mathcal{X} \to [0, 1]$, reward distributions (over $[0, 1]$) $M(x)$ such that $f(x)$ is the expectation of $M(x)$.

For each round $t = 1, 2, \ldots, n$;

1. The player chooses an arm $X_t \in \mathcal{X}$.
2. The environment draws the reward $Y_t$ from $M(X_t)$ (and independently from the past given $X_t$).

Goal: Maximize (in expectation) the cumulative rewards. Equivalently we want to minimize the cumulative regret

$$R_n = \mathbb{E} \sum_{t=1}^{n} \left( \max_{x \in \mathcal{X}} f(x) - Y_t \right).$$
Motivating examples

- **Calibrating the temperature** or levels of other inputs to a reaction so as to maximize the yield of a chemical process.
- Pricing a new product with uncertain demand in order to maximize revenue.
- In general: online parameter tuning of numerical methods.
- Note: in the pricing problem different product lines could also be tested while tuning the price ⇒ hybrid continuous/discrete set of arms.
Motivating examples

- **Calibrating the temperature** or levels of other inputs to a reaction so as to maximize the yield of a chemical process.
- **Pricing** a new product with uncertain demand in order to maximize revenue.
- In general: online parameter tuning of numerical methods.
- Note: in the pricing problem different product lines could also be tested while tuning the price ⇒ hybrid continuous/discrete set of arms.
Motivating examples

- **Calibrating the temperature** or levels of other inputs to a reaction so as to maximize the yield of a chemical process.
- **Pricing** a new product with uncertain demand in order to maximize revenue
- In general: online parameter tuning of numerical methods.

Note: in the pricing problem different product lines could also be tested while tuning the price ⇒ hybrid continuous/discrete set of arms.
Motivating examples

- **Calibrating the temperature** or levels of other inputs to a reaction so as to maximize the yield of a **chemical process**.
- **Pricing** a new product with **uncertain demand** in order to maximize revenue.
- In general: online parameter tuning of numerical methods.
- Note: in the pricing problem **different product lines** could also be tested while tuning the price ⇒ **hybrid continuous/discrete set of arms**.
Summary of the talk

- We present a new strategy, **Hierarchical Optimistic Optimization (HOO)**. It is based on a tree-representation of the search space, that we explore non-uniformly thanks to upper confidence bounds assigned to each nodes.

- Main theoretical result: if one knows the local regularity of the mean-payoff function around its maximum, then it is possible to obtain a cumulative regret of order $\sqrt{n}$.

- In particular, using $n$ (noisy) evaluation of the function we can find the maximum at a precision $1/\sqrt{n}$, independently of the ambient dimension! Note that in a minimax sense, one can only find the maximum at a precision $n^{-1}/(d+2)$.
Summary of the talk

- We present a new strategy, **Hierarchical Optimistic Optimization (HOO)**. It is based on a tree-representation of the search space, that we explore non-uniformly thanks to upper confidence bounds assigned to each nodes.

- Main theoretical result: if one knows the local regularity of the mean-payoff function around its maximum, then it is possible to obtain a cumulative regret of order $\sqrt{n}$.

- In particular, using $n$ (noisy) evaluation of the function we can find the maximum at a precision $1/\sqrt{n}$, independently of the ambient dimension! Note that in a minimax sense, one can only find the maximum at a precision $n^{-1/(d+2)}$. 
Summary of the talk

• We present a new strategy, **Hierarchical Optimistic Optimization (HOO)**. It is based on a **tree-representation** of the search space, that we explore non-uniformly thanks to **upper confidence bounds** assigned to each nodes.

• Main theoretical result: if one knows the **local regularity** of the mean-payoff **function around its maximum**, then it is possible to obtain a cumulative regret of order $\sqrt{n}$.

• In particular, using $n$ (noisy) evaluation of the function we can find the **maximum at a precision $1/\sqrt{n}$**, independently of the ambient dimension! Note that in a minimax sense, one can only find the maximum at a precision $n^{-1/(d+2)}$. 
Local regularity around the maximum

Let $\ell$ be *dissimilarity* measure, that is, a non-negative mapping $\ell : \mathcal{X}^2 \rightarrow \mathbb{R}$ satisfying $\ell(x, x) = 0$.

**Assumption (Weakly Lipschitz)**

For all $x \in \mathcal{X}$ and $\epsilon \geq 0$, if $x \in \mathcal{X}_\epsilon = \{x \in \mathcal{X}, f^* - f(x) \leq \epsilon\}$ then for any $y \in \mathcal{X}$, $f(x) - f(y) \leq \max(\epsilon, \ell(x, y))$. 

S. Bubeck & R. Munos & G. Stoltz & C. Szepesvari  
$\mathcal{X}$-Armed Bandits
Local regularity around the maximum

Let $\ell$ be *dissimilarity* measure, that is, a non-negative mapping $\ell : \mathcal{X}^2 \to \mathbb{R}$ satisfying $\ell(x, x) = 0$.

**Assumption (Weakly Lipschitz)**

For all $x \in \mathcal{X}$ and $\epsilon \geq 0$, if $x \in \mathcal{X}_\epsilon = \{ x \in \mathcal{X} : f^* - f(x) \leq \epsilon \}$ then for any $y \in \mathcal{X}$, $f(x) - f(y) \leq \max(\epsilon, \ell(x, y))$. 
Local regularity around the maximum

Let $\ell$ be *dissimilarity* measure, that is, a non-negative mapping $\ell : \mathcal{X}^2 \to \mathbb{R}$ satisfying $\ell(x, x) = 0$.

**Assumption (Weakly Lipschitz)**

For all $x \in \mathcal{X}$ and $\epsilon \geq 0$, if $x \in \mathcal{X}_\epsilon = \{x \in \mathcal{X}, f^* - f(x) \leq \epsilon\}$ then for any $y \in \mathcal{X}$, $f(x) - f(y) \leq \max(\epsilon, \ell(x, y))$. 
HOO receives as input a sequence \((\mathcal{P}_{h,i})_{h \geq 0, 1 \leq i \leq 2^h}\) of subsets of \(\mathcal{X}\) satisfying:

1. \(\mathcal{P}_{0,1} = \mathcal{X}\),
2. \(\mathcal{P}_{h,i} = \mathcal{P}_{h+1,2i-1} \cup \mathcal{P}_{h,2i}\),
3. \(\exists \rho \in (0, 1) : \text{diam}(\mathcal{P}_{h,i}) \leq \rho^h\) where \(\text{diam}(\mathcal{P}_{h,i}) = \sup_{x, y \in \mathcal{P}_{h,i}} \ell(x, y)\).

We view this as a tree where node \((h, i)\) (at depth \(h\) and position \(i\)) is associated to the domain \(\mathcal{P}_{h,i}\).
HOO - Input

- HOO receives as input a sequence \((\mathcal{P}_{h,i})_{h \geq 0, 1 \leq i \leq 2^h}\) of subsets of \(\mathcal{X}\) satisfying:

  1. \(\mathcal{P}_{0,1} = \mathcal{X}\),
  2. \(\mathcal{P}_{h,i} = \mathcal{P}_{h+1,2i-1} \cup \mathcal{P}_{h,2i}\).
  3. \(\exists \rho \in (0, 1) : \text{diam}(\mathcal{P}_{h,i}) \leq \rho^h\)

- We view this as a tree where node \((h, i)\) (at depth \(h\) and position \(i\)) is associated to the domain \(\mathcal{P}_{h,i}\).
HOO - Input

HOO receives as input a sequence \((\mathcal{P}_{h,i})_{h \geq 0, 1 \leq i \leq 2^h}\) of subsets of \(\mathcal{X}\) satisfying:

1. \(\mathcal{P}_{0,1} = \mathcal{X}\),
2. \(\mathcal{P}_{h,i} = \mathcal{P}_{h+1,2i-1} \cup \mathcal{P}_{h,2i}\).
3. \(\exists \rho \in (0, 1): \text{diam} (\mathcal{P}_{h,i}) \leq \rho^h\) where \(\text{diam} (\mathcal{P}_{h,i}) = \sup_{x,y \in \mathcal{P}_{h,i}} \ell(x,y)\).

We view this as a tree where node \((h, i)\) (at depth \(h\) and position \(i\)) is associated to the domain \(\mathcal{P}_{h,i}\).
HOO - Input

- HOO receives as input a sequence \( (\mathcal{P}_{h,i})_{h \geq 0, 1 \leq i \leq 2^h} \) of subsets of \( \mathcal{X} \) satisfying:
  1. \( \mathcal{P}_{0,1} = \mathcal{X} \),
  2. \( \mathcal{P}_{h,i} = \mathcal{P}_{h+1,2i-1} \cup \mathcal{P}_{h,2i} \).
  3. \( \exists \rho \in (0, 1) : \text{diam}(\mathcal{P}_{h,i}) \leq \rho^h \) where \( \text{diam}(\mathcal{P}_{h,i}) = \sup_{x,y \in \mathcal{P}_{h,i}} \ell(x, y) \).

- We view this as a tree where node \((h, i)\) (at depth \( h \) and position \( i \)) is associated to the domain \( \mathcal{P}_{h,i} \).
HOO - Input

- HOO receives as input a sequence \((P_{h,i})_{h,0,1\leq i\leq 2^h}\) of subsets of \(X\) satisfying:
  1. \(P_{0,1} = X\),
  2. \(P_{h,i} = P_{h+1,2i-1} \cup P_{h,2i}\).
  3. \(\exists \rho \in (0, 1) : \text{diam}(P_{h,i}) \leq \rho^h\) where
     \[ \text{diam}(P_{h,i}) = \sup_{x,y \in P_{h,i}} \ell(x, y). \]

- We view this as a tree where node \((h, i)\) (at depth \(h\) and position \(i\)) is associated to the domain \(P_{h,i}\).
HOO - Global strategy given $B$–values for each node

$B_{h,i}$

Turned-on nodes

Followed path

Selected node

Pulled point $X_t$
HOO - Definition of $B$–values

- Let $T_{h,i}(n)$ be the number of points we pulled in $(h, i)$.
- Let $\hat{\mu}_{h,i}(n)$ be the empirical average in the domain $(h, i)$.
- We consider the following upper confidence bound for each node already visited:

$$U_{h,i}(n) = \hat{\mu}_{h,i}(n) + \sqrt{\frac{2 \ln n}{T_{h,i}(n)} + \text{diam}(\mathcal{P}_{h,i})}.$$

- Our $B$–values are defined as:

$$B_{h,i}(n) = \min \left\{ U_{h,i}(n), \max\{ B_{h+1,2i-1}(n), B_{h+1,2i}(n) \} \right\}.$$
Let $T_{h,i}(n)$ be the number of points we pulled in $(h, i)$.

Let $\hat{\mu}_{h,i}(n)$ be the empirical average in the domain $(h, i)$.

We consider the following upper confidence bound for each node already visited:

$$U_{h,i}(n) = \hat{\mu}_{h,i}(n) + \sqrt{\frac{2\ln n}{T_{h,i}(n)}} + \text{diam}(\mathcal{P}_{h,i}).$$

Our $B$–values are defined as:

$$B_{h,i}(n) = \min \left\{ U_{h,i}(n), \max \{ B_{h+1, 2i-1}(n), B_{h+1, 2i}(n) \} \right\}.$$
HOO - Definition of $B$–values

- Let $T_{h,i}(n)$ be the number of points we pulled in $(h, i)$.
- Let $\hat{\mu}_{h,i}(n)$ be the empirical average in the domain $(h, i)$.
- We consider the following upper confidence bound for each node already visited:

\[
U_{h,i}(n) = \hat{\mu}_{h,i}(n) + \sqrt{\frac{2 \ln n}{T_{h,i}(n)}} + \text{diam}(P_{h,i}).
\]

- Our $B$–values are defined as:

\[
B_{h,i}(n) = \min \left\{ U_{h,i}(n), \max \left\{ B_{h+1,2i-1}(n), B_{h+1,2i}(n) \right\} \right\}.
\]
Let $T_{h,i}(n)$ be the number of points we pulled in $(h, i)$.

Let $\hat{\mu}_{h,i}(n)$ be the empirical average in the domain $(h, i)$.

We consider the following upper confidence bound for each node already visited:

$$U_{h,i}(n) = \hat{\mu}_{h,i}(n) + \sqrt{\frac{2 \ln n}{T_{h,i}(n)} + \text{diam}(P_{h,i})}.$$ 

Our B-values are defined as:

$$B_{h,i}(n) = \min \left\{ U_{h,i}(n), \max \{ B_{h+1,2i-1}(n), B_{h+1,2i}(n) \} \right\}.$$
HOO - Numerical Example

\[ n = 1000 \]

\[ n = 10000 \]
Main result

**Definition (Near-optimality dimension)**

Let $d \geq 0$ be such that $\mathcal{X}_\epsilon = \{x \in \mathcal{X}, f^* - f(x) \leq \epsilon \}$ can be packed with $O(\epsilon^{-d})$ balls of radius $\epsilon$. 
Main result

Definition (Near-optimality dimension)

Let $d \geq 0$ be such that $\mathcal{X}_\epsilon = \{x \in \mathcal{X}, f^* - f(x) \leq \epsilon\}$ can be packed with $O(\epsilon^{-d})$ balls of radius $\epsilon$. 
Main result

Definition (Near-optimality dimension)

Let \( d \geq 0 \) be such that \( \mathcal{X}_\epsilon = \{ x \in \mathcal{X}, f^* - f(x) \leq \epsilon \} \) can be packed with \( O(\epsilon^{-d}) \) balls of radius \( \epsilon \).

\[ \ell(x, y) = ||x - y|| \Rightarrow d = \frac{D}{2}. \]

\[ \ell(x, y) = ||x - y||^2 \Rightarrow d = 0. \]

Theorem

HOO satisfy \( R_n \leq \tilde{O}(n^{(d+1)/(d+2)}) \).
Main result

Definition (Near-optimality dimension)

Let $d \geq 0$ be such that $\mathcal{X}_\epsilon = \{ x \in \mathcal{X}, f^* - f(x) \leq \epsilon \}$ can be packed with $O(\epsilon^{-d})$ balls of radius $\epsilon$.

\[ \ell(x, y) = ||x - y|| \Rightarrow d = D/2. \]
\[ \ell(x, y) = ||x - y||^2 \Rightarrow d = 0. \]

Theorem

HOO satisfy $R_n \leq \tilde{O}(n^{(d+1)/(d+2)})$. 

S. Bubeck & R. Munos & G. Stoltz & C. Szepesvari

$\mathcal{X}$-Armed Bandits
Example

\[ \mathcal{X} = [0, 1]^D, \alpha \geq 0 \text{ and } f \text{ locally "} \alpha\text{-smooth" around (any of) its maximum } x^* \text{ (in finite number):} \]

\[ f(x^*) - f(x) = \Theta(||x - x^*||^{\alpha}) \text{ as } x \to x^*. \]

Theorem

Assume that we run HOO with diameters measured with
\[ \ell(x, y) = ||x - y||^\beta. \]

- Known smoothness: \( \beta = \alpha \). \( R_n \leq \tilde{O}(\sqrt{n}) \), i.e., the rate is independent of the dimension \( D \). Previously known for \( D = 1 \) or \( \alpha \leq 1 \).
- Smoothness underestimated: \( \beta < \alpha \).
  \( R_n \leq \tilde{O}(n^{(d+1)/(d+2)}) \) where \( d = D \left( \frac{1}{\beta} - \frac{1}{\alpha} \right) \).
- Smoothness overestimated: \( \beta > \alpha \). No guarantee. Note: UCT corresponds to \( \beta = +\infty \).
Example

$\mathcal{X} = [0, 1]^D$, $\alpha \geq 0$ and $f$ locally "$\alpha$-smooth" around (any of) its maximum $x^*$ (in finite number):

$$f(x^*) - f(x) = \Theta(||x - x^*||^\alpha) \text{ as } x \to x^*.$$

Theorem

Assume that we run HOO with diameters measured with

$\ell(x, y) = ||x - y||^\beta$.

- Known smoothness: $\beta = \alpha$. $R_n \leq \tilde{O}(\sqrt{n})$, i.e., the rate is independent of the dimension $D$. Previously known for $D = 1$ or $\alpha \leq 1$.
- Smoothness underestimated: $\beta < \alpha$. $R_n \leq \tilde{O}(n^{(d+1)/(d+2)})$ where $d = D \left(\frac{1}{\beta} - \frac{1}{\alpha}\right)$.
- Smoothness overestimated: $\beta > \alpha$. No guarantee. Note: UCT corresponds to $\beta = +\infty$. 

S. Bubeck & R. Munos & G. Stoltz & C. Szepesvari

$\mathcal{X}$-Armed Bandits
Example

\[ \mathcal{X} = [0, 1]^{D}, \quad \alpha \geq 0 \text{ and } f \text{ locally } "\alpha\text{-smooth}" \text{ around (any of) its maximum } x^* \text{ (in finite number):} \]

\[ f(x^*) - f(x) = \Theta(||x - x^*||^\alpha) \text{ as } x \to x^*. \]

Theorem

Assume that we run HOO with diameters measured with \( \ell(x, y) = ||x - y||^\beta \).

- **Known smoothness:** \( \beta = \alpha \). \( R_n \leq \tilde{O}(\sqrt{n}) \), i.e., the rate is independent of the dimension \( D \). Previously known for \( D = 1 \) or \( \alpha \leq 1 \).

- **Smoothness underestimated:** \( \beta < \alpha \). 
  \( R_n \leq \tilde{O}(n^{(d+1)/(d+2)}) \) where \( d = D \left( \frac{1}{\beta} - \frac{1}{\alpha} \right) \).

- **Smoothness overestimated:** \( \beta > \alpha \). No guarantee. Note: UCT corresponds to \( \beta = +\infty \).
Example

\[ \mathcal{X} = [0, 1]^D, \alpha \geq 0 \text{ and } f \text{ locally } "\alpha\text{-smooth}" \text{ around (any of) its maximum } x^* \text{ (in finite number):} \]

\[ f(x^*) - f(x) = \Theta(||x - x^*||^\alpha) \text{ as } x \to x*. \]

Theorem

Assume that we run HOO with diameters measured with

\[ \ell(x, y) = ||x - y||^\beta. \]

- **Known smoothness:** \( \beta = \alpha. \quad R_n \leq \tilde{O}(\sqrt{n}), \text{i.e., the rate is independent of the dimension } D. \) *Previously known for } D = 1 \text{ or } \alpha \leq 1.\]

- **Smoothness underestimated:** \( \beta < \alpha. \quad R_n \leq \tilde{O}(n^{(d+1)/(d+2)}) \text{ where } d = D \left( \frac{1}{\beta} - \frac{1}{\alpha} \right). \]

- **Smoothness overestimated:** \( \beta > \alpha. \) *No guarantee. Note: UCT corresponds to } \beta = +\infty.\]
Example

\[ X = [0, 1]^D, \quad \alpha \geq 0 \text{ and } f \text{ locally } "\alpha\text{-smooth}" \text{ around (any of) its maximum } x^* \text{ (in finite number):} \]

\[ f(x^*) - f(x) = \Theta(\|x - x^*\|^\alpha) \text{ as } x \rightarrow x^*. \]

Theorem

Assume that we run HOO with diameters measured with
\[ \ell(x, y) = \|x - y\|^{\beta}. \]

- **Known smoothness:** \( \beta = \alpha \). \( R_n \leq \tilde{O}(\sqrt{n}) \), i.e., the rate is independent of the dimension \( D \). Previously known for \( D = 1 \) or \( \alpha \leq 1 \).

- **Smoothness underestimated:** \( \beta < \alpha \).
  \( R_n \leq \tilde{O}(n^{(d+1)/(d+2)}) \) where \( d = D \left( \frac{1}{\beta} - \frac{1}{\alpha} \right) \).

- **Smoothness overestimated:** \( \beta > \alpha \). No guarantee. Note: UCT corresponds to \( \beta = +\infty \).