Influence of Monte-Carlo sampling on greedy algorithms

Mohammed-Raed Blel[1], Virginie Ehrlacher[1,2], Tony Lelièvre[1,2]

[1] Ecole des Ponts ParisTech

[2] INRIA

Journées IHP "Réduction de modèles et traitement géométrique des données", 9th November 2021

Outline of the talk

1. Greedy algorithms in Hilbert spaces: deterministic case

2. Greedy algorithms in Hilbert spaces: stochastic case
   - Variance reduction with reduced basis control variate
   - Monte-Carlo greedy algorithm

Outline of the talk

1. Greedy algorithms in Hilbert spaces: deterministic case

2. Greedy algorithms in Hilbert spaces: stochastic case
   - Variance reduction with reduced basis control variate
   - Monte-Carlo greedy algorithm

Solution set

Let V be a Hilbert space and $\mathcal{P} \subset \mathbb{R}^p$ be a set of parameter values for some $p \in \mathbb{N}^*$.

For all $\mu \in \mathcal{P}$, let $u_\mu \in V$ so that $\mathcal{M} := \{u_\mu, \; \mu \in \mathcal{P}\}$ is a compact subset of V.

For all $X \subset V$ finite-dimensional subspace of V, let Π_X denote the orthogonal projector of V onto X.

Example: For all $\mu \in \mathcal{P}$, $u_\mu \in V$ is the unique solution of a PDE system of the form

$$\mathcal{A}(u_\mu; \mu) = 0,$$

where $\mathcal{A}(\cdot; \mu)$ is a differential operator depending on the value of the parameters μ.

Definition

For all $n \in \mathbb{N}^*$,

$$d_n(\mathcal{M})_V := \inf_{\substack{V_n \subset V \\ \dim V_n = n}} \sup_{\mu \in \mathcal{M}} \|u_\mu - \Pi_{V_n} u_\mu\|_V$$

the Kolmogorov n-width of the set $\mathcal{M}$.

Question: How to find a constructive way to construct a sequence of linear spaces $(X_n)_{n\geq 1}$ which are quasi-optimal in the sense that

$$\sigma_n(\mathcal{M}; X_n)_V := \sup_{\mu \in \mathcal{P}} \|u_\mu - \Pi_{X_n} u_\mu\|_V$$

decays at a rate similar to $(d_n(\mathcal{M})_V)_{n\geq 1}$?

Greedy algorithm for reduced bases

- Initialization $n = 1$: Let $\mu_1 \in \mathcal{P}$ such that

$$\mu_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \|u_\mu\|_V^2.$$

Let $\theta_1 := \frac{u_{\mu_1}}{\|u_{\mu_1}\|_V}$ and $X_1 := \mathrm{Vect}\{\theta_1\} = \mathrm{Vect}\{u_{\mu_1}\}$.

- Iteration $n \geq 2$: Let $\mu_n \in \mathcal{P}$ such that

$$\mu_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \left\| u_\mu - \Pi_{X_{n-1}} u_\mu \right\|_V^2.$$

Let $\theta_n := \frac{u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n}}{\left\| u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n} \right\|_V}$ and $X_n := \mathrm{Vect}\{\theta_1, \cdots, \theta_n\} = \mathrm{Vect}\{u_{\mu_1}, \cdots, u_{\mu_n}\}$.

Soit $0 < \gamma \leq 1$.

- Initialization $n = 1$: Let $\mu_1 \in \mathcal{P}$ such that

$$\|u_{\mu_1}\|_V \geq \gamma \sup_{\mu \in \mathcal{P}} \|u_\mu\|_V.$$

Let $\theta_1 := \frac{u_{\mu_1}}{\|u_{\mu_1}\|_V}$ and $X_1 := \mathrm{Vect}\{\theta_1\} = \mathrm{Vect}\{u_{\mu_1}\}$.

- Iteration $n \geq 2$: Let $\mu_n \in \mathcal{P}$ such that

$$\|u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n}\|_V \geq \gamma \sup_{\mu \in \mathcal{P}} \|u_\mu - \Pi_{X_{n-1}} u_\mu\|_V.$$

Let $\theta_n := \frac{u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n}}{\left\|u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n}\right\|_V}$ and $X_n := \mathrm{Vect}\{\theta_1, \cdots, \theta_n\} = \mathrm{Vect}\{u_{\mu_1}, \cdots, u_{\mu_n}\}$.

Convergence rates of greedy algorithms

[Buffa, Maday, Patera, Prud'homme, Turinici, 2012]

[Binev,Cohen, Dahmen, DeVore, Petrova,Wojtaszczyk, 2011]

[DeVore, Petrova, Wojtaszczyk, 2013]

Theorem

- *For all* $n \in \mathbb{N}^*$,
$$\sigma_{2n}(\mathcal{M}; X_{2n})_V \leq \sqrt{2}\gamma^{-1}\sqrt{d_n(\mathcal{M})_V}.$$
- *If* $d_n(\mathcal{M})_V \leq Cn^{-\alpha}$, *then* $\sigma_n(\mathcal{M}; X_n)_V \leq \widetilde{C}n^{-\alpha}$.
- *If* $d_n(\mathcal{M})_V \leq Ce^{-cn^{\alpha}}$, *then* $\sigma_n(\mathcal{M}; X_n)_V \leq \widetilde{C}e^{-\widetilde{c}n^{\alpha}}$.

Reduced basis method:

- Offline stage: Select $\mu_1, \cdots, \mu_n \in \mathcal{P}$ using a (weak) greedy algorithm.
- Online stage: For all $\mu \in \mathcal{P}$, approximate u_μ by a linear combination of $u_{\mu_1}, \cdots, u_{\mu_n}$ (with a Galerkin method for instance).

Outline of the talk

1. Greedy algorithms in Hilbert spaces: deterministic case

2. Greedy algorithms in Hilbert spaces: stochastic case
   - Variance reduction with reduced basis control variate
   - Monte-Carlo greedy algorithm

Previous works related to greedy algorithms in a stochastic setting

- Smetana-Zahm-Patera (2018), Balabanov-Nouy (2019): Randomized residual-based error estimators
- Cohen-Dahmen-Devore-Nichols (2020): Random training sets $\mathcal{P}_{\text{train}} \subset \mathcal{P}$.
- Cohen-Dolbeault (2021): Sampling numbers
- Boyaval-Lelièvre... (2010): Variance reduction using a reduced basis control variate

Outline of the talk

1. Greedy algorithms in Hilbert spaces: deterministic case

2. Greedy algorithms in Hilbert spaces: stochastic case
   - Variance reduction with reduced basis control variate
   - Monte-Carlo greedy algorithm

Notation and framework

- Let $d \in \mathbb{N}^*$ and $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$.
- Let Z be a random vector of dimension d with law characterized by a probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$.
- Let $L^2_\nu(\mathbb{R}^d)$ be the set of functions $u : \mathbb{R}^d \to \mathbb{R}$ such that

$$\int_{\mathbb{R}^d} |u(z)|^2 \, d\nu(z) = \mathbb{E}\left[|u(Z)|^2\right] < +\infty.$$

For all $\mu \in \mathcal{P}$, let $u_\mu \in L^2_\nu(\mathbb{R}^d)$ so that $\mathcal{M} := \{u_\mu, \ \mu \in \mathcal{P}\}$ is a compact subset $L^2_\nu(\mathbb{R}^d)$.

Additional assumption: $\mathcal{M} \subset \mathcal{C}(\mathbb{R}^d)$.

Objective: Analysis of a variance reduction method proposed [Boyaval, Lelièvre et al., 2010] in order to efficiently compute

$$\mathbb{E}\left[u_\mu(Z)\right]$$

for a large number of values of $\mu \in \mathcal{P}$.

Notation

- Let $M \in \mathbb{N}^*$ and let $\overline{Z} := (Z_k)_{1 \leq k \leq M}$ a set of M iid randdom vectors distributed according to the law of Z.
- For all $u, v \in L^2_\nu(\mathbb{R}^d) \cap \mathcal{C}(\mathbb{R}^d)$, we denote by

$$\mathbb{E}_{\overline{Z}}(u) := \frac{1}{M} \sum_{k=1}^{M} u(Z_k),$$

$$\mathrm{Cov}_{\overline{Z}}(u, v) := \mathbb{E}_{\overline{Z}}(uv) - \mathbb{E}_{\overline{Z}}(u)\mathbb{E}_{\overline{Z}}(v),$$

$$\mathrm{Var}_{\overline{Z}}(u) := \sqrt{\mathrm{Cov}_{\overline{Z}}(u, u)} = \sqrt{\mathbb{E}_{\overline{Z}}(u^2) - \mathbb{E}_{\overline{Z}}(u)^2}.$$

Standard Monte-Carlo method

- Let $M_{\mathrm{ref}} \in \mathbb{N}^*$ (so that M_{ref} is typically very large), and let $\overline{Z}^{\mathrm{ref}} := \left(Z_k^{\mathrm{ref}}\right)_{1 \leq k \leq M_{\mathrm{ref}}}$ be a set of M_{ref} iid randdom vectors distributed according to the law of Z.

- The **standard Monte-Carlo method** consists in computing, for all $\mu \in \mathcal{P}$, an approximation of $\mathbb{E}\left[u_\mu(Z)\right]$ as the empirical mean

$$\mathbb{E}_{\overline{Z}^{\mathrm{ref}}}\left(u_\mu\right) = \frac{1}{M_{\mathrm{ref}}} \sum_{k=1}^{M_{\mathrm{ref}}} u_\mu\left(Z_k^{\mathrm{ref}}\right).$$

- The **statistical error** given by this method is of the order of

$$\frac{\sqrt{\mathrm{Var}\left[u_\mu(Z)\right]}}{\sqrt{M_{\mathrm{ref}}}} \leq \frac{C}{\sqrt{M_{\mathrm{ref}}}} \quad \text{où } C := \sup_{\mu \in \mathcal{P}} \sqrt{\mathrm{Var}\left[u_\mu(Z)\right]}.$$

For all $\mu \in \mathcal{P}$, the computational cost scales like $\mathcal{O}(M_{\mathrm{ref}})$.

Reduced basis control variate

- **Idea:** Construct (in an online stage) an approximation of $u_\mu(Z)$ as a linear combination of $u_{\mu_1}(Z), \cdots, u_{\mu_n}(Z)$ for some well-chosen $\mu_1, \cdots, \mu_n \in \mathcal{P}$ (the choice being done in an offline stage).
- **Offline stage:** Selection of $n \in \mathbb{N}^*$, $\mu_1, \cdots, \mu_n \in \mathcal{P}$ and computation of

$$\mathbb{E}_{\overline{Z}^{\text{ref}}}\left(u_{\mu_i}\right) \quad \forall 1 \leq i \leq n.$$

- Let $M_{\text{small}} \in \mathbb{N}^*$ (so that $M_{\text{small}} \ll M_{\text{ref}}$), and let $\overline{Z}^{\text{small}} := \left(Z_k^{\text{small}}\right)_{1 \leq k \leq M_{\text{small}}}$ be a set of M_{small} iid randdom vectors distributed according to the law of Z (and independent of $\overline{Z}^{\text{ref}}$).
- Computation of

$$\mathbb{E}_{\overline{Z}^{\text{small}}}\left(u_{\mu_i}\right) \quad \text{and} \quad A_{ij} := \text{Cov}_{\overline{Z}^{\text{small}}}\left(u_{\mu_i}, u_{\mu_j}\right) \quad \forall 1 \leq i, j \leq n.$$

Reduced basis control variate

- **Online stage:** For all $\mu \in \mathcal{P}$, computation of

$$B_i := \mathrm{Cov}_{\overline{Z}^{\mathrm{small}}}\left(u_{\mu_i}, u_\mu\right) \quad \forall 1 \leq i \leq n \qquad \mathcal{O}(nM_{\mathrm{small}}).$$

- Let $\Lambda^\mu := (\lambda_i^\mu)_{1\leq i\leq n} \in \mathbb{R}^n$ solution of

$$\Lambda^\mu = \operatorname*{argmin}_{\Lambda:=(\lambda_i)_{1\leq i\leq n}\in\mathbb{R}^n} \mathrm{Var}_{\overline{Z}^{\mathrm{small}}}\left(u_\mu - \sum_{i=1}^n \lambda_i u_{\mu_i}\right), \tag{1}$$

or equivalently, of

$$A\Lambda^\mu = B, \quad B = (B_i)_{1\leq i\leq n}, \; A = (A_{ij})_{1\leq i,j\leq n}. \qquad \mathcal{O}(n^2)$$

- Compute an approximation of
$\mathbb{E}\left[u_\mu(Z)\right] = \mathbb{E}\left[\left(u_\mu - \sum_{i=1}^n u_{\mu_i}\right)(Z)\right] + \sum_{i=1}^n \lambda_i^\mu \mathbb{E}\left[u_{\mu_i}(Z)\right]$ as

$$\mathbb{E}_{\overline{Z}^{\mathrm{small}}}\left(u_\mu - \sum_{i=1}^n \lambda_i^\mu u_{\mu_i}\right) + \sum_{i=1}^n \lambda_i^\mu \mathbb{E}_{\overline{Z}^{\mathrm{ref}}}\left(u_{\mu_i}\right). \qquad \mathcal{O}(M_{\mathrm{small}} + n)$$

Reduced basis control variate

- Better (in terms of a computational complexity) than a standard Monte Carlo approach if
$$nM_{\text{small}} \ll M_{\text{ref}}.$$

- **What about accuracy?** The statistical error is then (assuming that $|\lambda_i^\mu| \leq \kappa$ for some $\kappa > 0$ for the sake of simplicity) of the order of
$$\frac{\sqrt{\mathrm{Var}\left[\left(u_\mu - \sum_{i=1}^n \lambda_i^\mu u_{\mu_i}\right)(Z)\right]}}{\sqrt{M_{\text{small}}}} + \frac{C\kappa n}{\sqrt{M_{\text{ref}}}}.$$

  The statistical error will then be comparable to a standard Monte-Carlo method if
$$\mathrm{Var}\left[\left(u_\mu - \sum_{i=1}^n \lambda_i^\mu u_{\mu_i}\right)(Z)\right] \approx \frac{M_{\text{small}}}{M_{\text{ref}}}.$$

  It is then natural to choose $(\lambda_\mu^i)_{1\leq i\leq n}$ such that $\mathrm{Var}\left[\left(u_\mu - \sum_{i=1}^n \lambda_i^\mu u_{\mu_i}\right)(Z)\right]$ is as small as possible, hence (1).

Outline of the talk

1. Greedy algorithms in Hilbert spaces: deterministic case

2. Greedy algorithms in Hilbert spaces: stochastic case
   - Variance reduction with reduced basis control variate
   - Monte-Carlo greedy algorithm

Offline stage: choice of $\mu_1, \cdots, \mu_n \in \mathcal{P}$ with a greedy algorithm with Monte-Carlo sampling

- **Important remark:**
  The space $V := L^2_\nu(\mathbb{R}^d)/\mathbb{R}$ (the set of functions of $L^2_\nu(\mathbb{R}^d)$ defined up to an additive constant) is a Hilbert space embedded with the scalar product:

$$\forall u, v \in V, \quad \langle u, v\rangle_V := \mathrm{Cov}[u(Z), v(Z)] := \mathbb{E}[u(Z)v(Z)] - \mathbb{E}[u(Z)]\mathbb{E}[v(Z)].$$

  The associated norm is then

$$\|u\|_V^2 := \mathrm{Var}[u(Z)] = \mathrm{Cov}[u(Z), u(Z)] = \mathbb{E}[u(Z)^2] - \mathbb{E}[u(Z)]^2.$$

- For any finite-dimensional subspace $X \subset V$, and for all $u \in V$, $\Pi_X u$ is the unique element of V such that

$$\mathrm{Var}[u(Z) - \Pi_X u(Z)] = \min_{v \in X} \mathrm{Var}[u(Z) - v(Z)].$$

Ideal greedy algorithm

- Initialization $n = 1$: Let $\mu_1 \in \mathcal{P}$ such that

$$\mu_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \|u_\mu\|_V^2.$$

Let $\theta_1 := \frac{u_{\mu_1}}{\|u_{\mu_1}\|_V}$ and $X_1 := \mathrm{Vect}\{\theta_1\} = \mathrm{Vect}\{u_{\mu_1}\}$.

- Iteration $n \geq 2$: Let $\mu_n \in \mathcal{P}$ such that

$$\mu_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \left\| u_\mu - \Pi_{X_{n-1}} u_\mu \right\|_V^2.$$

Let $\theta_n := \frac{u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n}}{\left\| u_{\mu_n} - \Pi_{X_{n-1}} u_{\mu_n} \right\|_V}$ and $X_n := \mathrm{Vect}\{\theta_1, \cdots, \theta_n\} = \mathrm{Vect}\{u_{\mu_1}, \cdots, u_{\mu_n}\}$.

Ideal greedy algorithm

- Initialization $n = 1$: Let $\mu_1 \in \mathcal{P}$ such that

$$\mu_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}[u_\mu(Z)].$$

Let $\theta_1 := \frac{u_{\mu_1}}{\sqrt{\mathrm{Var}[u_{\mu_1}(Z)]}}$ and $X_1 := \mathrm{Vect}\{\theta_1\} = \mathrm{Vect}\{u_{\mu_1}\}$.

- Iteration $n \geq 2$: Let $\mu_n \in \mathcal{P}$ such that

$$\mu_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}[u_\mu(Z) - \Pi_{X_{n-1}} u_\mu(Z)].$$

Let $\theta_n := \frac{u_{\mu n} - \Pi_{X_{n-1}} u_{\mu n}}{\sqrt{\mathrm{Var}[u_{\mu n}(Z) - \Pi_{X_{n-1}} u_{\mu n}(Z)]}}$ and $X_n := \mathrm{Vect}\{\theta_1, \cdots, \theta_n\} = \mathrm{Vect}\{u_{\mu_1}, \cdots, u_{\mu_n}\}$.

Monte Carlo greedy algorithm

- Exact variances cannot be computed in practice, only **empirical variances**.
- For all $n \in \mathbb{N}^*$, let $M_n \in \mathbb{N}^*$, $\overline{Z}^n := (Z_k^n)_{1 \leq k \leq n}$ iid randdom vectors distributed according to the law of Z and independent of Z.

  Let us also assume that $m, n \in \mathbb{N}^*$ and that for all $1 \leq k \leq M_n$ and $1 \leq l \leq M_m$, Z_k^n is independent of Z_l^m as soon as $n \neq m$ or $k \neq l$.

- For all $X \subset V$ and all $u \in V$, let $\overline{\Pi}_X^n u \in V$ be the unique solution to

$$\mathrm{Var}_{\overline{Z}^n}\left(u - \overline{\Pi}_X^n u\right) = \inf_{v \in X} \mathrm{Var}_{\overline{Z}^n}(u - v).$$

Monte-Carlo greedy algorithm

- Initialization $n = 1$: Let $\overline{\mu}_1 \in \mathcal{M}$ such that
$$\overline{\mu}_1 \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^1}(u_\mu).$$
Let $\overline{\theta}_1 := \frac{u_{\overline{\mu}_1}}{\sqrt{\mathrm{Var}[u_{\overline{\mu}_1}(Z)]}}$ and $\overline{X}_1 := \mathrm{Vect}\{\overline{\theta}_1\} = \mathrm{Vect}\{u_{\overline{\mu}_1}\}$.
- Iteration $n \geq 2$: Let $\overline{\mu}_n \in \mathcal{P}$ such that
$$\overline{\mu}_n \in \operatorname*{argmax}_{\mu \in \mathcal{P}} \mathrm{Var}_{\overline{Z}^n}\left(u_\mu - \overline{\Pi}^n_{\overline{X}_{n-1}} u_\mu\right).$$
Let $\overline{\theta}_n := \frac{u_{\overline{\mu}_n} - \Pi_{\overline{X}_{n-1}} u_{\overline{\mu}_n}}{\sqrt{\mathrm{Var}[u_{\overline{\mu}_n}(Z) - \Pi_{\overline{X}^{n-1}} u_{\overline{\mu}_n}(Z)]}}$ and
$\overline{X}_n := \mathrm{Vect}\{\overline{\theta}_1, \cdots, \overline{\theta}_n\} = \mathrm{Vect}\{u_{\overline{\mu}_1}, \cdots, u_{\overline{\mu}_n}\}$.

Question: under which assumptions can one guarantee that the Monte-Carlo greedy algorithm is a weak greedy algorithm with parameter $\gamma \in (0, 1]$ with high probability?

In particular, how to choose the sequence $(M_n)_{n\in\mathbb{N}^*}$?

Assumption on ν

- **Assumption (A):** The probability measure ν is such that there exists $\alpha > 1$ an $\beta > 0$ such that

$$\int_{\mathbb{R}} e^{\beta |x|^{\alpha}} \, d\nu(x) < +\infty.$$

- **Consequence:** We can use results from [Fournier,Guilin,2015] about concentration inequalities of the empirical measure of ν in Wasserstein distance.

Concentration inequality in Wasserstein distance

- Let $\mathcal{L}$ be the set of Lipschitz functions on $\mathbb{R}^d$, and for all $u \in \mathcal{L}$, let $\|u\|_{\mathcal{L}}$ its Lipschitz constant.
- Let $\phi : \mathbb{R}_+^* \to \mathbb{R}_+^*$ defined by

$$\forall \kappa \in \mathbb{R}_+^*, \quad \phi(\kappa) := \begin{cases} \kappa^2 \mathbb{1}_{\kappa \leq 1} + \kappa^\alpha \mathbb{1}_{\kappa > 1} & \text{if } d = 1, \\ (\kappa / \log(2 + 1/\kappa)^2) \mathbb{1}_{\kappa \leq 1} + \kappa^\alpha \mathbb{1}_{\kappa > 1} & \text{if } d = 2, \\ \kappa^d \mathbb{1}_{\kappa \leq 1} + \kappa^\alpha \mathbb{1}_{\kappa > 1} & \text{if } d \geq 3. \end{cases} \tag{2}$$

Theorem (Fournier, Guilin, 2015)

Assume that ν satisfies assumption (A). Then, there exist $c, C > 0$ (which only depend on ν, d, α and β), such that for all $M \in \mathbb{N}^*$, $\overline{Z} := (Z_k)_{1 \leq k \leq M}$ iid random vectors distributed according to ν and for all $\kappa > 0$,

$$\mathbb{P}\left[\mathcal{T}_1\left(\overline{Z}\right) \geq \kappa\right] \leq C e^{-cM\phi(\kappa)},$$

where

$$\mathcal{T}_1\left(\overline{Z}\right) := \sup_{f \in \mathcal{L}; \|f\|_{\mathcal{L}} \leq 1} \left|\mathbb{E}[f(Z)] - \mathbb{E}_{\overline{Z}}(f)\right|.$$

Assumption on $\mathcal{M}$

$$\mathcal{M} := \{u_\mu,\ \mu \in \mathcal{M}\}$$

Assumption (B): The set $\mathcal{M}$ satisfies the four conditions:

(B1) $\mathcal{M}$ is a compact subset of V and let $C_2 := \sup_{\mu \in \mathcal{P}} \sqrt{\mathrm{Var}\left[u_\mu(Z)\right]} < \infty$;

(B2) $\mathcal{M} \subset \mathcal{L}$ and $C_{\mathcal{L}} := \sup_{\mu \in \mathcal{P}} \|u_\mu\|_{\mathcal{L}} < +\infty$;

(B3) $\mathcal{M} \subset L^\infty(\mathbb{R}^d)$ and $C_\infty := \sup_{\mu \in \mathcal{P}} \|u_\mu - \mathbb{E}[u_\mu(Z)]\|_{L^\infty} < +\infty$;

(B4) for all $n \in \mathbb{N}^*$, $d_n(\mathcal{M})_V > 0$.

Notation

For all $n \in \mathbb{N}^*$, let

$$\overline{Z}^{1:n} = \left(\overline{Z}^1, \cdots, \overline{Z}^n\right)$$

$$\overline{\sigma}_{n-1} = \max_{\mu \in \mathcal{P}} \sqrt{\mathrm{Var}_{\overline{Z}^n}\left(u_\mu - \overline{\Pi}^n_{\overline{X}_{n-1}} u_\mu\right)},$$

$$\widehat{\sigma}_{n-1} = \max_{\mu \in \mathcal{P}} \sqrt{\mathrm{Var}\left[u_\mu(Z) - \Pi_{\overline{X}_{n-1}} u_\mu(Z) \,\middle|\, \overline{Z}^{1:(n-1)}\right]}.$$

$$C^n_\infty := \max\left(C_\infty, \|\overline{\theta}_1 - \mathbb{E}[\overline{\theta}_1(Z)]\|_{L^\infty}, \cdots, \|\overline{\theta}_n - \mathbb{E}[\overline{\theta}_n(Z)]\|_{L^\infty}\right)$$

$$C^n_{\mathcal{L}} := \max\left(C_{\mathcal{L}}, \|\overline{\theta}_1\|_{\mathcal{L}}, \cdots, \|\overline{\theta}_n\|_{\mathcal{L}}\right).$$

$$\kappa_0 := \frac{\left(1 - \gamma^2\right)\widehat{\sigma}_0^2}{8 C_\infty C_{\mathcal{L}}}; \tag{3}$$

$$\forall n \geq 1, \quad \kappa_n := \frac{\min\left(\frac{1}{2n}, \quad \frac{(1-\gamma^2)\widehat{\sigma}_n^2}{(n+1)\left(9C_2^2+4\right)}\right)}{6 C^n_\infty C^n_{\mathcal{L}}}. \tag{4}$$

Remark: For all $n \in \mathbb{N}$, κ_n is a deterministic function of $\overline{Z}^{1:n}$.

Theorem (Blel, Ehrlacher, Lelièvre, 2021)

Let $0 < \delta < 1$ *and* $(\delta_n)_{n\in\mathbb{N}^*} \subset (0,1)^{\mathbb{N}^*}$ *a sequence such that* $\Pi_{n\in\mathbb{N}^*}(1-\delta_n) \geq 1-\delta$*.*
Let us assume that $\mathcal{M}$ *satisfies assumption (B) and that* ν *satisfies assumption (A).*
Let us assume that there exists $0 < \gamma < 1$ *such that for all* $n \in \mathbb{N}^*$*,* $M_n \in \mathbb{N}^*$ *is a* $\overline{Z}^{1:(n-1)}$*-measurable random variable which satisfies almost surely:*

$$\forall n \geq 1, \quad M_n \geq -\ln\left(\frac{\delta_n}{C}\right)\frac{1}{c\phi\left(\kappa_{n-1}\right)}, \tag{5}$$

Then, the Monte-Carlo greedy algorithm is a weak greedy algorithm with parameter γ *with probability at least* $1-\delta$*.*

Conclusion

Summary:
Theoretical bounds on the number of samples to be used at each iteration of the Monte-Carlo greedy algorithm so that it can be theoretically guaranteed to be a weak greedy algorithm with high probability.

Open questions:

- Pessimistic bounds: better inequalities? For now, heuristic strategies to choose the sequence $(M_n)_{n \geq 1}$ in practice.
- Independent uniform sampling: what about adaptive sampling?
- Z is assumed to be a finite dimensional random vector: what happens if Z is infinite-dimensional? (Brownian motion)

Thank you for your attention!