Optimal designs for discrete least-squares approximation

Cécile Haberstich[1], **Anthony Nouy**[2], **Guillaume Perrin**[3]

ETICS2021 | 13 - 17 septembre 2021

[1] CEA,DAM,DIF, F-91297, Arpajon, France, cecile.haberstich@cea.fr
[2] Centrale Nantes, LMJL, UMR CNRS 6629, France
[3] COSYS, Université Gustave Eiffel, 77420 Champs-sur-Marne, France

Outline

1. Boosted optimal weighted least-squares projection
   - Optimal weighted least-squares
   - Boosted optimal weighted least-squares method.
   - Optimal sampling for L2 approximation.

2. Towards a new criteria to choose the set of samples.
   - Support points
   - Illustration in dimension 1
   - Illustration in higher dimension

Boosted optimal weighted least-squares projection

Objective

- *Let u be a function defined on $\mathcal{X} \subset \mathbb{R}^d$,*

$$\begin{cases} u : \mathcal{X} \to \mathbb{R} \\ \quad x \mapsto y = u(x) \end{cases}$$

- *Consider u in $L^2_\mu(\mathcal{X})$ the Hilbert space of square-integrable real-valued functions defined on $\mathcal{X} \subset \mathbb{R}^d$.*
- *We want to **construct $u^\star$ an approximation of u**, using samples $\{x^i, y^i = u(x^i)\}_{i=1}^n$*

→ *with **controlled error** $\|u - u^\star\|^2 \leq \varepsilon$,*

→ *while using only a **few evaluations** n.*

- *Propose a projection method onto a **linear space** $V_m \subset L^2_\mu(\mathcal{X})$ which is stable and whose construction requires a number of evaluations n close to $\dim(V_m) = m$.*

Least-squares methods

- The best approximation of u by an element of V_m is given by the **orthogonal projection** :

$$P_{V_m} u = \arg\min_{v \in V_m} \|u - v\|^2_{L^2_\mu}.$$

- It is not computable in practice → replaced by a weighted least-squares projection :

$$Q_{V_m} u = \arg\min_{v \in V_m} \frac{1}{n} \sum_{i=1}^{n} w(x^i)(v(x^i) - u(x^i))^2 \text{ where } x^i \sim \rho,$$

where $w \geq 0$ is a weight function and ρ is a measure.

Stability of least-squares methods

- $\{\varphi_j\}_{j=1}^m$ is a given **orthonormal** basis of V_m.
- The **stability of** Q_{V_m} is measured by the properties of the **empirical Gram matrix** $\hat{\boldsymbol{G}}_n$.
- The empirical Gram matrix $\hat{\boldsymbol{G}}_n$ associated to a sample $\{x^i\}_{i=1}^n$ is given by

$$(\hat{\boldsymbol{G}}_n)_{k,l} = \frac{1}{n}\sum_{i=1}^n w(x^i)\varphi_k(x^i)\varphi_l(x^i).$$

- **More stability** if $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$ **closer to 0**.
- How to choose the sampling density ρ and weights to have $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$ close to 0 while using few evaluations n ?

Optimal least-squares methods [Cohen and Migliorati., 2017]

Theorem 1 (Optimal weighted least-squares)

Let $d\rho(x) = w(x)^{-1} d\mu(x)$ *with* $w(x)^{-1} = \frac{1}{m}\sum_{j=1}^{m} \varphi_j(x)^2$.
Let $\eta \in (0,1)$ *and* $\delta \in (0,1)$, *and* $x^1, \cdots, x^n$ *be i.i.d from* ρ. *For* $n \geq \delta^{-2} m \log(2m\eta^{-1})$, *it holds*

$$\mathbb{P}(\|\hat{G}_n - I\| \leq \delta) \geq 1 - \eta.$$

The approximation $Q^C_{V_m} u$ *defined by* $Q_{V_m} u$ *if* $\|\hat{G}_n - I\| < \delta$ *and* 0 *otherwise satisfies*

$$\mathbb{E}(\|u - Q^C_{V_m} u\|^2) \leq (1-\delta)^{-1} \|u - P_{V_m} u\|^2 + \eta \|u\|^2.$$

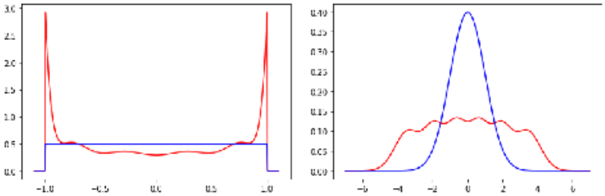


FIGURE – Illustration of the optimal measures associated with the uniform and the gaussian distributions (with $m = 6$).

Optimal least-squares methods [Cohen and Migliorati., 2017]

Theorem 1 (Optimal weighted least-squares)

Let $d\rho(x) = w(x)^{-1} d\mu(x)$ *with* $w(x)^{-1} = \frac{1}{m}\sum_{j=1}^{m} \varphi_j(x)^2$.
Let $\eta \in (0,1)$ *and* $\delta \in (0,1)$, *and* $x^1, \cdots, x^n$ *be i.i.d from* ρ. *For* $n \geq \delta^{-2} m \log(2m\eta^{-1})$, *it holds*

$$\mathbb{P}(\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| \leq \delta) \geq 1 - \eta.$$

The approximation $Q^C_{V_m} u$ *defined by* $Q_{V_m} u$ *if* $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| < \delta$ *and* 0 *otherwise satisfies*

$$\mathbb{E}(\|u - Q^C_{V_m} u\|^2) \leq (1-\delta)^{-1} \|u - P_{V_m} u\|^2 + \eta \|u\|^2.$$

Comments

→ *Improving stability (smaller* δ*) and the chance to have this stability (smaller* η*) implies higher* n.

→ *Even if* δ *is close to* 1, n *may be high compared to an interpolation method* $(n = m)$.

Next, we propose another measure $\tilde{\rho}$ based on ρ to improve the properties of $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$.

Outline

1. Boosted optimal weighted least-squares projection
   - Optimal weighted least-squares
   - Boosted optimal weighted least-squares method.
   - Optimal sampling for L2 approximation.
2. Towards a new criteria to choose the set of samples.
   - Support points
   - Illustration in dimension 1
   - Illustration in higher dimension

Boosted optimal least-squares method (BLS)

1. **Resampling :** draw M independent n-samples $\{\boldsymbol{x}^{n,i}\}_{i=1}^{M}$, with $\boldsymbol{x}^{n,i} = (x^{1,i}, \cdots, x^{n,i})$, where all $x^{j,i} \sim \rho$ are i.i.d. and choose the one which minimizes $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$.

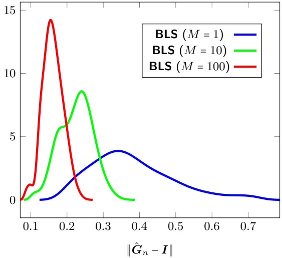


FIGURE – Distribution of $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\|$ for $\delta = 0.9$

Interpretations

→ *Resampling improves the stability for a given probability η.*

2. **Conditioning by rejection :** Repeat step 1 until $\|\hat{\boldsymbol{G}}_n - \boldsymbol{I}\| < \delta$ → output sample $\widetilde{\boldsymbol{x}} = (\widetilde{x}^1, \cdots, \widetilde{x}^n)$. This ensures stability almost surely.

Boosted optimal least-squares method (BLS)

3. **Greedy removal of samples :** Let $\hat{G}_K$ be the empirical Gram matrix associated with sample $\{\tilde{x}^i : i \in K\}$. Begin with $K = \{1, \cdots, n\}$ and while $\|\hat{G}_K - I\| \leq \delta$ successively remove a sample $\{k^\star\}$ that minimizes $\|\hat{G}_{K \setminus \{k^\star\}} - I\|$.

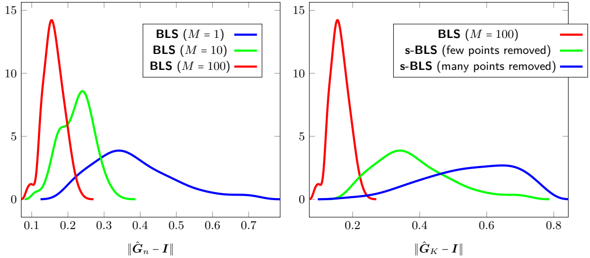


FIGURE – Distribution of $\|\hat{G}_K - I\|$ for $\delta = 0.9$

We do not have any theoretical guarantees to say that we can remove points !

Boosted optimal least-squares (BLS)

Theorem 2 (Control of the error bound)

Let $\eta \in (0,1)$ and $\delta \in (0,1)$, and let $Q_{V_m} u$ be the boosted optimal least-squares projection with $n \geq \delta^{-2} m \log(2m\eta^{-1})$ and $\#K \geq n_0$.
The error of approximation is bounded in expectation

$$\mathbb{E}(\|u - Q_{V_m} u\|^2) \leq C \|u - P_{V_m} u\|^2$$

with $C = (1 + \frac{n}{n_0}(1-\delta)^{-1}(1-\eta^M)^{-1}M)$.
Further assuming that u is uniformly bounded, we can obtain a better bound.

Comments

→ *If $n_0 = \frac{n}{\beta}$, for some $\beta \geq 1$ → quasi-optimality property (in expectation).*

→ *C increases with M and the number of removed points.*

Boosted optimal least-squares distribution (example)

- Empirical boosted least-squares distribution is represented by the histograms.
- Dashed lines represent the optimal distribution [Cohen and Migliorati., 2017].

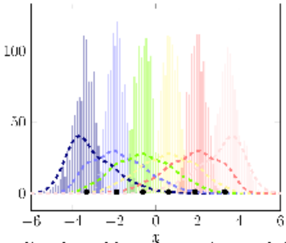


FIGURE – Gaussian distribution, Hermite polynomial approximation spaces.

Comments

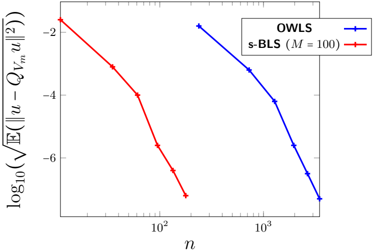
→ *Boosted distribution is peaked.*

→ *Gauss-Hermite quadrature points are represented in black. They may be aligned with the modes of the boosted distribution.*

Illustration on a simple example

$$u(x) = \frac{1}{(1 - \frac{0.5}{2d}\sum_{i=1}^d x_i)^{d+1}}$$ defined on $\mathcal{X} = [-1,1]^d$, equipped with the uniform measure

- Hyperbolic cross polynomial approximation spaces with Legendre polynomials for different m with $d = 2$.



- Guaranteed stability with probability greater than 0.99 for the OWLS method and almost surely for the s-BLS method.

Comments

→ *With subsampling (OWLS → s-BLS) n is significantly decreased.*

Limits of boosted least-squares... :(

Comments

→ *Theoretical bound C is high compared to what we observe in numerical experiments.*

→ *The **offline cost** is important (compared to an interpolation method for example). It is due to two reasons → generate M times a n-samples*
→ greedy strategy (even with the approximate technique).

Open questions :

- Theoretical guarantees for the **convergence** of the greedy subsampling ?
- Improve the error bound (maybe with other sets of points) ?
- Improve the offline cost (in this case with other sets of points) ?

Outline

1 Boosted optimal weighted least-squares projection
- Optimal weighted least-squares
- Boosted optimal weighted least-squares method.
- Optimal sampling for L2 approximation.

2 Towards a new criteria to choose the set of samples.
- Support points
- Illustration in dimension 1
- Illustration in higher dimension

There is hope … Recent works [Cohen and Dolbeault, 2021]

Theorem 3 (Optimal sampling for L2 approx [Cohen and Dolbeault, 2021])

For some universal constants $\widetilde{C}, \widetilde{\beta}$*, and for any* m*-dimensional space* $V_m \subset L^2$*, there exists a random sampling* $\boldsymbol{x} = x^1, \cdots, x^n$*, with* $n \leq \widetilde{\beta} m$ *such that for any* $u \in L^2$*, for* $Q_{V_m} u = \arg\min_{v \in V_m} \sum_{i=1}^m w(x^i) |u(x^i) - v(x^i)|$,

$$\mathbb{E}(\|u - Q_{V_m} u\|^2) \leq \widetilde{C} \|u - P_{V_m} u\|^2$$

Comments

→ *Intuitive justification for boosted least-squares (subsampling step makes sense).*

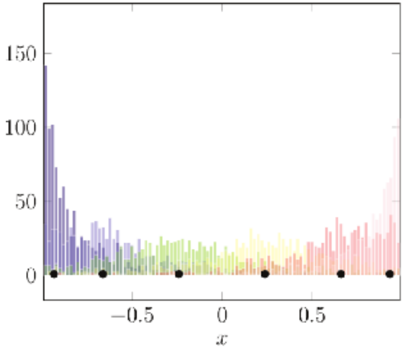
→ *Open questions : construct the random sampling, find thigher bounds for the constants ?*

Outline

1. Boosted optimal weighted least-squares projection
   - Optimal weighted least-squares
   - Boosted optimal weighted least-squares method.
   - Optimal sampling for L2 approximation.

2. Towards a new criteria to choose the set of samples.
   - Support points
   - Illustration in dimension 1
   - Illustration in higher dimension

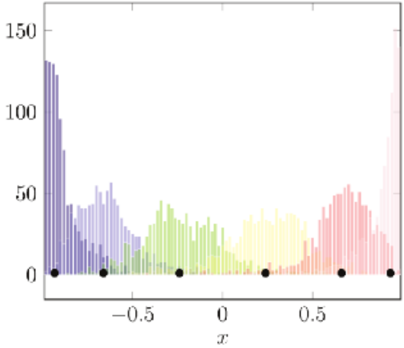
Observations and intuitions



(a) $M = 1$

- Distributions of the $x^{(i)}, i = 1, \cdots, 6$ with x^6 sampled from the boosted method for $V_m = \mathbb{P}_5$ and μ the uniform measure.
- In black, are represented the Gauss-Legendre nodes.

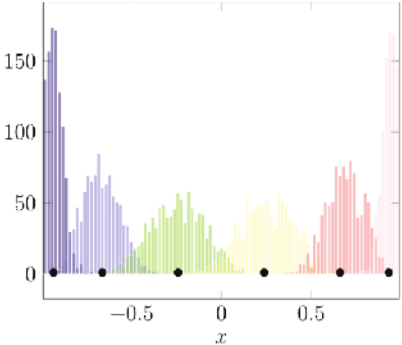
Observations and intuitions



(b) $M = 10$

- Distributions of the $x^{(i)}, i = 1, \cdots, 6$ with x^6 sampled from the boosted method for $V_m = \mathbb{P}_5$ and μ the uniform measure.
- In black, are represented the Gauss-Legendre nodes.

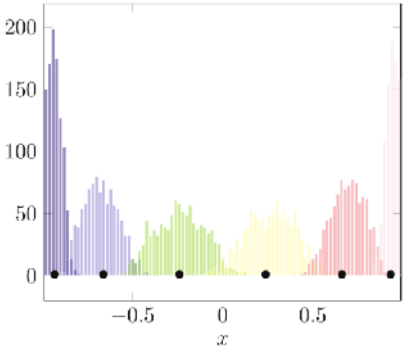
Observations and intuitions



(c) $M = 50$

- Distributions of the $x^{(i)}, i = 1, \cdots, 6$ with $\boldsymbol{x}^6$ sampled from the boosted method for $V_m = \mathbb{P}_5$ and μ the uniform measure.
- In black, are represented the Gauss-Legendre nodes.

Observations and intuitions



(d) $M = 100$

- The more repetitions, the more concentrate are these distributions around the black nodes.
- Can we summarize the distribution with a set of points ?

Support points [Mak and Joseph, 2018]

Support points

- New way to compact a continuous probability distribution ρ into a set of **representative points**.
- "Support points can be viewed as **optimal sampling points in ρ** (in the sense of maximum energy) for any desired sample size n".

Theoretical guarantees

- Support points converge in distribution to ρ.
- Points obtained by **minimizing an energy distance** $\rightarrow$ quite efficient in high-dimensions.
- There exist a lot of theoretical results...

Computation

- **Easy to compute** (just need to sample a lot of points from the distribution ρ.)

Energy distance between two distributions

Definition 4 (Energy distance)

Let ρ and μ be two probability distributions. Let $\boldsymbol{X}, \boldsymbol{X}' \sim \rho$ and $\boldsymbol{Y}, \boldsymbol{Y}' \sim \mu$. The energy distance between ρ and μ is defined as

$$E(\rho, \mu) = 2\mathbb{E}\|\boldsymbol{X} - \boldsymbol{Y}\|_2 - \mathbb{E}\|\boldsymbol{X} - \boldsymbol{X}'\|_2 - \mathbb{E}\|\boldsymbol{Y} - \boldsymbol{Y}'\|_2$$

Also, when $\mu = \boldsymbol{\rho}_n$, is the e.d.f for $\{x_i\}_{i=1}^n$, this energy distance becomes :

$$E(\rho, \boldsymbol{\rho}_n) = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[\|x_i - \boldsymbol{Y}\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2 - \mathbb{E}\|\boldsymbol{Y} - \boldsymbol{Y}'\|$$

Definition 5

Let ρ be a probability distribution. Let $\boldsymbol{Y} \sim \rho$. For a fixed points set size n, the support points of ρ are defined as :

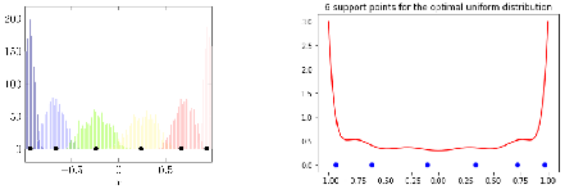
$$\{x_i\}_{i=1}^n \in \arg\min_{x_1,\cdots,x_n} E(\rho, \boldsymbol{\rho}_n)$$

$$\{x_i\}_{i=1}^n \in \arg\min_{x_1,\cdots,x_n} \{\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[\|x_i - \boldsymbol{Y}\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2\}$$

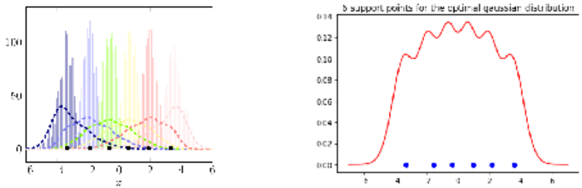
Outline

1. Boosted optimal weighted least-squares projection
   - Optimal weighted least-squares
   - Boosted optimal weighted least-squares method.
   - Optimal sampling for L2 approximation.
2. Towards a new criteria to choose the set of samples.
   - Support points
   - Illustration in dimension 1
   - Illustration in higher dimension

Support points of the optimal distributions



Histograms of the boosted distribution, black dots : quadrature points, in red optimal weighted distribution and in blue its 6 support points.

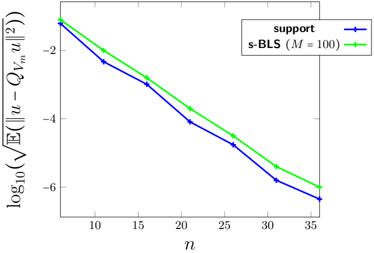


Histograms of the boosted distribution, black dots : quadrature points, in red optimal weighted distribution and in blue its 6 support points.

Support points of the optimal distributions

$u(x) = \dfrac{1}{(1+5x^2)}$ defined on $\mathcal{X} = [-1,1]$, equipped with the uniform measure

- Polynomial approximation spaces with Legendre polynomials for different m.



- Given cost $(n = m)$ for the s-BLS method.

Comments

→ *In dimension 1, approximation error is slightly better with the support points.*

→ *In dimension 1, the support points are equal to the point set with minimal L_2-discrepancy.*

→ *But in general in dimension 1, things are quite simple.*

Outline

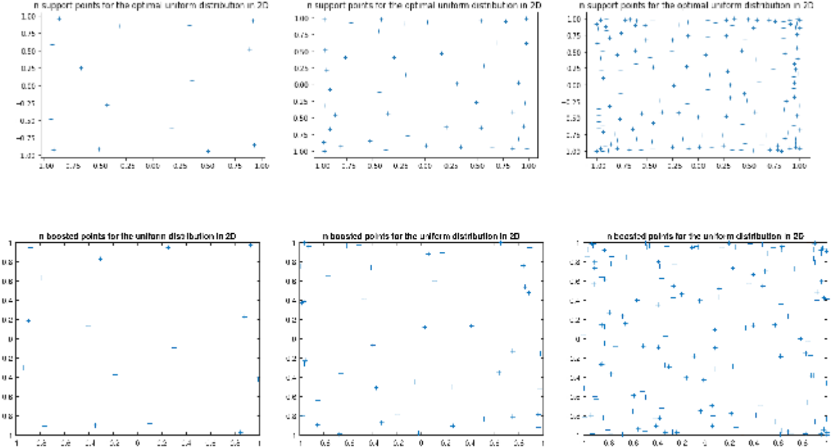
1 Boosted optimal weighted least-squares projection

- Optimal weighted least-squares
- Boosted optimal weighted least-squares method.
- Optimal sampling for L2 approximation.

2 Towards a new criteria to choose the set of samples.

- Support points
- Illustration in dimension 1
- Illustration in higher dimension

Support points with the optimal distributions in higher dimensions

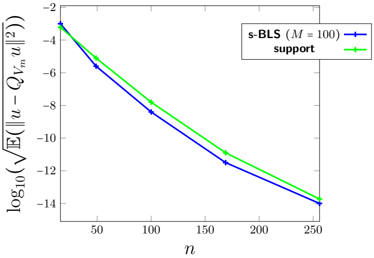


$n = m$ $\qquad$ $n = 3m$ $\qquad$ $n = 10m$

Illustration on a first example

$$u(x) = \frac{1}{(1 - \frac{0.5}{2d}\sum_{i=1}^{d} x_i)^{d+1}}$$ defined on $\mathcal{X} = [-1,1]^d$, equipped with the uniform measure

- Full polynomial approximation spaces with Legendre polynomials for different m with $d = 2$.



- $n = m$ for both methods

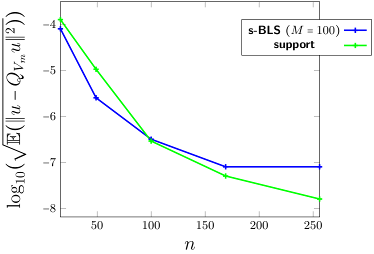
Comments

→ *In the case where $n = m$, the boosted least-squares are slightly better than approx. with support points.*

Illustration on a second example

$$u(x_1, x_2) = \frac{1}{(10 + x_1 + 0.5x_2)^2}$$ defined on $\mathcal{X} = \mathbb{R}^2$, equipped with the gaussian measure

- Full polynomial approximation spaces with Hermite polynomials for different m with $d = 2$.



- $n = m$ for both methods.

Comments

→ *In the case where $n = m$, the support points are better than the boosted least-squares for higher values of m.*

Work perspectives

- Theoretical analysis is clearly not straightforward.
- **Koksma-Hlawaka inequality** → upper bound on the integration error I :

$$I(g,\rho,\rho_n) = \left| \int_{\mathcal{X}} g(x)d\rho(x) - \sum_{i=1}^{n} g(x^i) \right| \leq \left\| \frac{\partial^p g}{\partial \boldsymbol{x}} \right\|_{L_q} D_r(\rho,\rho_n) \text{ with } \frac{1}{q} + \frac{1}{r} = 1$$

→ How to bound the term $D_r(\rho, \rho_n)$?

- **[Mak and Joseph, 2018]** with hypotheses on ρ and g, propose a bound in

$$\mathcal{O}\left(\frac{1}{n^{1/2}}(\log n)^{-(\alpha-1)/2}\right) \text{ with } \alpha > 1$$

.

Work perspectives

- In **[Migliorati and Nobile, 2015]** discrete least-squares approximation on multivariate polynomial space with evaluations at low-discrepancy point sets is stable and accurate if $n \approx m^2 \log(m)$.
- Talk by Chris. J. Oates this week.

- **[Gruber, 2004]** gives the minimum error of numerical integration formulae for classes of Hölder continuous functions and optimum sets of nodes → but asymptotic results $(n \to \infty)$.
- (In **[Pagès, 1998]**, with an optimal grid for the quadratic quantization and if u is $\mathcal{C}^1$ with Lipschitz continuous differential Du

$$\mathbb{E}(\|u - Q_{V_m} u\|^2) \leq \mathcal{O}\left(\frac{1}{n^{2/d}}\right)$$

Conclusions

- Recent works from [Cohen and Dolbeault, 2021] provide some **intuitive justification** to the boosted least-squares methods.
- However, it remains some offline computational issues.
- We use the **support points** from [Mak and Joseph, 2018] combined with the optimal measure from [Cohen and Migliorati., 2017] to propose a **new design for $L2$ approximation**, that is very fast to generate (in particular compared to BLS).
- On numerical experiments we see that

→ for $d = 1$, this is very efficient.

→ for $d > 1$, this is competitive with boosted least-squares.

- But **theoretical analysis is not straightforward**.
- For now, it seems that **more hypotheses on the functions** are necessary (compared to the boosted least-squares).

References I

Cohen, A. and Dolbeault, M. (2021).
Optimal pointwise sampling for l2 approximation.

Cohen, A. and Migliorati., G. (2017).
Optimal weighted least-squares methods.
SMAI Journal of Computational Mathematics, 86(3) :181–203.

Gruber, P. M. (2004).
Optimum quantization and its applications.
Advances in Mathematics, 186(2) :456–497.

Haberstich, C., Nouy, A., and Perrin, G. (2020).
Boosted optimal weighted least-squares methods.
arXiv :1912.07075.

Mak, S. and Joseph, V. R. (2018).
Support points.

References II

Migliorati, G. and Nobile, F. (2015).
Analysis of discrete least squares on multivariate polynomial spaces with evaluations at low-discrepancy point sets.
Journal of Complexity, 31(4) :517–542.

Pagès, G. (1998).
A space quantization method for numerical integration.
Journal of Computational and Applied Mathematics, 89(1) :1–38.

Thank you for your attention.
Do you have any questions ?