Weighted least-squares for randomised L^2 approximation

Albert Cohen, **Matthieu Dolbeault**

Laboratoire Jacques-Louis Lions (LJLL), Sorbonne Université

March 10, 2022

Setting

Let $D \subset \mathbb{R}^d$ domain, $F \subset L^2(D,\mu)$ set of functions on D

Goal

Approximate $f \in F$ based on point values at $x_1, \dots, x_m \in D$

Sampling numbers

$$g_m(F, L^2) = \inf_{x_1,\dots,x_m \in D} \inf_{\varphi_1,\dots,\varphi_m \in L^2} \sup_{f \in F} \left\| f - \sum_{i=1}^{m} f(x_i)\varphi_i \right\|_{L^2}$$

Approximation numbers

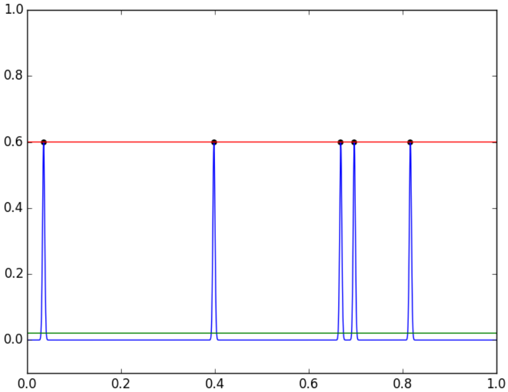
$$a_n(F, L^2) = \inf_{L_1,\dots,L_n : H \to \mathbb{C}} \inf_{\varphi_1,\dots,\varphi_n \in L^2} \sup_{f \in F} \left\| f - \sum_{i=1}^{n} L_i(f)\varphi_i \right\|_{L^2}$$

Point evaluations are not continuous in L^2

Take $V_n = \text{Span}(\varphi_1, \ldots, \varphi_n)$ a subspace of L^2

$$F = \{f \in L^2, d(f, V_n) \leqslant \varepsilon\}$$

Then $a_n(F, L^2) = \varepsilon$ but $g_n(F, L^2) = \infty$



Relaxed problems

- Compare $g_m(F, L^2)$ to $a_n(F, L^\infty)$
  I. Limonova and V.N. Temlyakov, *On sampling discretization in L^2* (2020)
  V. N. Temlyakov, *On optimal recovery in L^2*, JoC (2020)

Relaxed problems

- Compare $g_m(F, L^2)$ to $a_n(F, L^\infty)$
  I. Limonova and V.N. Temlyakov, *On sampling discretization in L^2* (2020)
  V. N. Temlyakov, *On optimal recovery in L^2*, JoC (2020)
- Compare $g_m(F, L^2)$ to $\sqrt{\frac{1}{n}\sum_{k>n} a_k(F, L^2)^2}$
  L. Kaemmerer, T. Ullrich, and T. Volkmer *Worst case recovery guarantees for least squares approximation using random samples*, CA (2019)
  N. Nagel, M. Schäfer, and T. Ullrich, *A new upper bound for sampling numbers*, FoCM (2020)
  D. Krieg and M. Ullrich, *Function values are enough for L^2-approximation*, FoCM (2020) and JoC (2021)

Relaxed problems

- Compare $g_m(F, L^2)$ to $a_n(F, L^\infty)$
  I. Limonova and V.N. Temlyakov, *On sampling discretization in L^2* (2020)
  V. N. Temlyakov, *On optimal recovery in L^2*, JoC (2020)
- Compare $g_m(F, L^2)$ to $\sqrt{\frac{1}{n}\sum_{k>n} a_k(F, L^2)^2}$
  L. Kaemmerer, T. Ullrich, and T. Volkmer *Worst case recovery guarantees for least squares approximation using random samples*, CA (2019)
  N. Nagel, M. Schäfer, and T. Ullrich, *A new upper bound for sampling numbers*, FoCM (2020)
  D. Krieg and M. Ullrich, *Function values are enough for L^2-approximation*, FoCM (2020) and JoC (2021)
- Compare the expected error g_m^{ran} over random points to $a_n(F, L^2)$

Expected L^2 error

Randomized sampling numbers

$$g_m^{\mathrm{ran}}(F, L^2) = \inf_{\sigma} \inf_{\varphi : D^m \times \mathbb{C}^m \to V_n} \sup_{f \in F} \sqrt{\mathbb{E}_{(x_1,\dots,x_m)\sim\sigma} \left\| f - \varphi(x_i, f(x_i)) \right\|_{L^2}^2}$$

If $\sigma_i \ll \mu$, $f(x_i)$ is almost surely well defined.

Framing

$$a_n(F, L^2) \leqslant g_n^{\mathrm{ran}}(F, L^2) \leqslant g_n(F, L^2)$$

Theorem (M.D. and A. Cohen, *Optimal pointwise sampling for L^2 approximation*, JoC 2022)

There exist universal constants $C, K > 0$ such that

$$g_m^{\mathrm{ran}}(F, L^2) \leqslant K a_n(F, L^2)$$

with $m \leqslant Cn$

Random sampling

Take $(b_1, \dots, b_n)$ an orthonormal basis of the optimal V_n

Christoffel function

$$\rho(x) = \frac{1}{n} \sum_{j=1}^{n} |b_j(x)|^2$$

Sample $x_1, \dots, x_m$ i.i.d according to the probability measure $\rho \, d\mu$
Define weights $w_i = 1/\sqrt{\rho(x_i)}$, a discretisation $N : f \mapsto (w_i f(x_i))_{i \leqslant m}$, and

$$G = (w_i b_j(x_i))_{i \leqslant m, j \leqslant n} \in \mathbb{C}^{m \times n}$$

Weighted least-squares approximation

$$Af := b \cdot G^+ Nf = b \cdot (G^* G)^{-1} G^* Nf$$

Sketch of the proof

Denote P the L^2-orthogonal projection onto V_n, and $\bar{f} = f - Pf$, then assuming that

$$\|G^+\|_{2\to 2}^2 \leqslant \frac{K_1}{m} \quad \text{and} \quad \frac{1}{m}\mathbb{E}\|N\bar{f}\|_2^2 \leqslant K_2\|\bar{f}\|_{L^2}^2$$

gives

$$\begin{aligned}
\mathbb{E}\|f - Af\|_{L^2}^2 &= \|f - Pf\|_{L^2}^2 + \mathbb{E}\|Af - Pf\|_{L^2}^2 \\
&= \|\bar{f}\|_{L^2}^2 + \|A\bar{f}\|_{L^2}^2 \\
&\leqslant a_n^2 + \mathbb{E}\|G^+\|_{2\to 2}^2\|N\bar{f}\|_2^2 \\
&\leqslant a_n^2 + \frac{K_1}{m}\mathbb{E}\|N\bar{f}\|_2^2 \\
&\leqslant (1 + K_1K_2)a_n^2
\end{aligned}$$

Satisfying the two conditions

By choice of the weights and sampling measure

$$\frac{1}{m}\mathbb{E}\|N\bar{f}\|_2^2 = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}w_i^2|\bar{f}(x_i)|^2 = \int_D \frac{1}{\rho}|\bar{f}|^2\rho\, d\mu = \|\bar{f}\|_{L^2}^2$$

Moreover $\|G^+\|_{2\to 2}^2 = s_{\min}(G)^{-2} = \lambda_{\min}(G^*G)^{-1}$ and

$$\mathbb{E}(G^*G) = \sum_{i=1}^{m}\mathbb{E}y_i^* y_i = m\left(\int_D \frac{1}{\rho} b_j b_k \rho\, d\mu\right)_{j,k} = m\mathrm{I}$$

Matrix Chernoff bound

Theorem (R. Ahlswede and A. Winter, see J. Tropp, *User-Friendly tail bounds for sums of random matrices*, FoCM 2012)

Let

$$\Lambda = \frac{1}{m} \sum_{i=1}^{m} y_i^* y_i \in \mathbb{C}^{n \times n}$$

with (y_i) i.i.d vectors, such that $\mathbb{E}(\Lambda) = \mathrm{I}$ and $\|y_i\|_2^2 \leqslant \delta$. Then

$$\mathbb{P}(\|\Lambda - \mathrm{I}\,\|_{2 \to 2} > 1/2) \leqslant 2n e^{-m/10\delta}$$

Here $\Lambda = \frac{1}{m} G^* G$ so $y_i = w_i \left(b_j(x_i)\right)_j$ satisfies the hypotheses

$$\|y_i\|_2^2 = \frac{1}{\rho(x_i)} \sum_{j=1}^{n} |b_j(x_i)|^2 = n =: \delta$$

Consequence : for $m \geqslant 10 n \log(4n)$, the event $\mathcal{E} = \{\Lambda \geqslant \frac{1}{2}\,\mathrm{I}\}$ occurs with probability at least $\frac{1}{2}$

Resampling

We resample $x_1, \ldots, x_m$ until $\mathcal{E}$ happens. In the end

$$G^* G = m\Lambda \geqslant \frac{m}{2} \mathrm{I}$$

and

$$\frac{1}{m}\mathbb{E}(\|N\bar{f}\|_2^2|\mathcal{E}) = \frac{1}{m}\frac{\mathbb{E}(\|N\bar{f}\|_2^2)}{\mathbb{P}(\mathcal{E})} \leqslant \frac{2}{m}\mathbb{E}\|N\bar{f}\|_2^2 = 2\|\bar{f}\|_{L^2}^2.$$

Now the two conditions hold, with $m = \mathcal{O}(n \log n)$

A. Cohen and G. Migliorati, *Optimal weighted least squares methods,* SMAI JCM (2017)
C. Haberstich, A. Nouy, and G. Perrin, *Boosted optimal weighted least-squares* (2019)

Kadison-Singer problem / Weaver's theorem

Theorem (A. Marcus, D. Spielman and N. Srivastava, *Interlacing families II*, AoM 2015)

Let $y_1, \dots, y_m \in \mathbb{C}^n$ such that $\|y_i\|_2^2 \leqslant \delta$ and $\sum_{i=1}^m y_i^* y_i = \mathrm{I}$. Then there exists a partition $S_1 \sqcup S_2$ of $\{1, \dots, m\}$ such that

$$\sum_{i \in S_j} y_i^* y_i \leqslant \frac{(1+\sqrt{2\delta})^2}{2} \mathrm{I}, \quad j = 1, 2$$

Corollary (S. Nitzan, A. Olevskii and A. Ulanovskii, *Exponential frames on unbounded sets*, Proc. AMS, 2016)

Let $y_1, \dots, y_m \in \mathbb{C}^n$ such that $\|y_i\|_2^2 \leqslant \delta$ and $\alpha \, \mathrm{I} \leqslant \sum_{i=1}^m y_i^* y_i \leqslant \beta \, \mathrm{I}$, with $0 < \delta \leqslant \alpha < \beta$. Then there exists a partition $S_1 \sqcup S_2$ of $\{1, \dots, m\}$ such that

$$\frac{1 - 5\sqrt{\delta/\alpha}}{2} \alpha \, \mathrm{I} \leqslant \sum_{i \in S_j} y_i^* y_i \leqslant \frac{1 + 5\sqrt{\delta/\alpha}}{2} \beta \, \mathrm{I}, \quad j = 1, 2$$

Application

Idea : Iteratively split the sample S into S_1, S_2, and keep S_j with probability $p_j = |S_j|/|S|$
The previous lemma garantees to preserve

$$G^* G \geqslant K_1 m \,\mathrm{I}$$

Moreover

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{|S_j|}\|N\bar{f}\|_{S_j}^2\right) &= \mathbb{E}_S\left(\frac{p_1}{|S_1|}\|N\bar{f}\|_{S_1}^2 + \frac{p_2}{|S_2|}\|N\bar{f}\|_{S_2}^2\right) \\
&= \mathbb{E}\left(\frac{1}{|S|}\|N\bar{f}\|_S^2\right) \\
&= \dots \\
&= \mathbb{E}_{\{1,\dots,m\}}\frac{1}{m}\|N\bar{f}\|_2^2 \leqslant 2\|\bar{f}\|_{L^2}^2 = K_2\|\bar{f}\|_{L^2}^2
\end{aligned}$$

Reproducing Kernel Hilbert Spaces

Let $H(K) \subset L^2(D,\mu)$ be a separable RKHS with a kernel K of finite trace

$$\int_D K(x,x)d\mu(x) < \infty$$

There exists an L^2-orthonormal family $(b_n)_{n\geqslant 0}$ such that $(a_n b_n)_{n\geqslant 0}$ is orthonormal in H and

$$K(x,y) = \sum_{n\geqslant 0} |a_n|^2 \overline{b_n(x)} b_n(y)$$

almost everywhere. We take

$$F = \{f \in H, \|f\|_H \leqslant 1\}$$

Previous results

- D. Krieg and M. Ullrich ; L. Kämmerer, T. Ullrich, and T. Volkmer :

$$g_n^2 \leqslant C \frac{\log n}{n} \sum_{k \geqslant \lfloor cn/\log n \rfloor} a_k^2$$

- N. Nagel, M. Schäfer, and T. Ullrich :

$$g_n^2 \leqslant C \frac{\log n}{n} \sum_{k \geqslant \lfloor cn \rfloor} a_k^2$$

- D. Krieg and M. Ullrich : Generalisation to arbitrary Banach classes F, but with $\|(a_k)\|_{\ell^p}$ for $p < 2$
- A. Hinrichs, D. Krieg, E. Novak and J. Vybiral : For any non-negative and non-increasing sequence $a \in \ell^2(\mathbb{N})$, there exists a RKHS H such that $(a_k)_{k \in \mathbb{N}} = a$ and

$$g_n^2 \geqslant \frac{1}{8n} \sum_{k \geqslant n} a_k^2$$

for infinitely many values of n

A new bound

Theorem

There exist universal constants $C, c > 0$ such that for all $n \geqslant 1$

$$g_n^2 \leqslant \frac{C}{n} \sum_{k \geqslant \lfloor cn \rfloor} a_k^2$$

Corollary

If $a_n \lesssim n^{-\alpha} \log^\beta n$ for $\alpha > \frac{1}{2}$ and $\beta \in \mathbb{R}$, then $g_n \lesssim n^{-\alpha} \log^\beta n$

If $a_n \lesssim n^{-1/2} \log^\beta n$ for $\alpha = \frac{1}{2}$ and $\beta < -\frac{1}{2}$, then $g_n \lesssim n^{-1/2} \log^{\beta+1/2} n$

Rescaling to remove the density

Sampling density

$$\rho_n = \frac{1}{2}\left(\frac{1}{n}\sum_{k<n}|b_k(x)|^2 + \frac{\sum_{k\geqslant n} a_k^2|b_k(x)|^2}{\sum_{k\geqslant n} a_k^2}\right)$$

Change of variables

$$\widetilde{K}(x,y) = \frac{K(x,y)}{\sqrt{\rho_n(x)}\sqrt{\rho_n(y)}}, \quad d\tilde{\mu} = \rho_n\, d\mu$$

$$\widetilde{H} = \left\{\frac{f}{\sqrt{\rho_n}}, f \in H\right\}, \quad \|g\|_{\widetilde{H}} = \|\sqrt{\rho_n} g\|_H$$

WLOG we can assume that $\rho_n = 1$

Random sample

Draw i.i.d. random points $x_1, \dots, x_m \in D$ according to μ, and define

$$(y_i)_k = \begin{cases} b_k(x_i) & \text{if } k < n \\ \alpha\, a_k b_k(x_i) & \text{if } k \geqslant n \end{cases}, \quad \alpha := \left(a_n^2 + \frac{1}{m} \sum_{k \geqslant n} a_k^2 \right)^{-1/2}$$

Then

$$\|y_i\|_2^2 = \sum_{k<n} |\eta_k(x_i)|^2 + \alpha^2 \sum_{k \geqslant n} |g_k(x_i)|^2 \leqslant 2n\rho_n(x_i) = 2n$$

and

$$\mathbb{E}(y_i y_i^*) = \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & \alpha^2 a_n^2 & \\ 0 & & & \ddots \end{pmatrix} =: \Lambda$$

Concentration inequality for infinite matrices

Theorem (S. Mendelson and A. Pajor, see M. Moëller and T. Ullrich or N. Nagel, M. Schäfer and T. Ullrich)

Let $y_1, \dots, y_m$ be i.i.d. random sequences from $\ell^2(\mathbb{N})$ satisfying $\|y_i\|_2 \leqslant 2n$ almost surely and $\mathbb{E}(y_i y_i^*) = \Lambda$ with $\|\Lambda\|_{2\to 2} \leqslant 1$. Then

$$\mathbb{P}\left(\left\| \frac{1}{m} \sum_{i=1}^{m} y_i y_i^* - \Lambda \right\|_{2\to 2} > t \right) \leqslant 2^{3/4} m \exp\left(-\frac{mt^2}{42n} \right)$$

For $m \geqslant Cn \log n$, there exists a sample $x_1, \dots, x_m$ such that

$$\left\| \frac{1}{m} \sum_{i=1}^{m} y_i y_i^* - \Lambda \right\|_{2\to 2} \leqslant \frac{1}{2}$$

Change of basis

Complete

$$V_n = \mathrm{Span}\{b_0, \dots, b_{n-1}\}$$

into an L^2-orthogonal basis $(\hat{b}_j)_{j<p} = (b_0, \dots, b_{n-1}, \hat{b}_n, \dots, \hat{b}_{p-1})$ of

$$V_p := V_n \oplus \mathrm{Span}\left\{ \sum_{k=0}^{\infty} (y_i)_k b_k, 1 \leqslant i \leqslant m \right\}$$

Then

$$\hat{b} = (\hat{b}_j)_{j<p} = U\, b = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & U' \end{pmatrix} \begin{pmatrix} (b_k)_{k<n} \\ (b_k)_{k\geqslant n} \end{pmatrix}$$

so

$$\hat{\Lambda} = U \Lambda U^* = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & \hat{\Lambda}' \end{pmatrix}, \quad \|\hat{\Lambda}'\|_{2\to 2} \leqslant \alpha^2 a_n^2$$

Adding artificial rank-one matrices

Decompose

$$\mathrm{I} - \Lambda = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{I} - \Lambda' \end{pmatrix} = \sum_{j=1}^{p-n} z_j z_j^*$$

and take

$$y_i = \sqrt{\frac{m}{m_{j(i)}}}\, z_{j(i)}, \quad m_j = \left\lceil \frac{m}{2n} \|z_j\|_2^2 \right\rceil$$

with $j(i) \in \{1, \ldots, p-n\}$ such that $\{y_i,\ i = m+1, \ldots, q\}$ contains exactly m_j copies of each z_j/m_j. In this way

$$\|y_i\|_2^2 \leqslant m \|z_{j(i)}\|^2 m_{j(i)}^{-2} \leqslant 2n$$

and

$$\frac{1}{m} \sum_{i=m+1}^{q} y_i y_i^* = \sum_{j=1}^{p-n} \frac{m_j}{n_j} z_j z_j^* = I - \Lambda$$

so

$$\left\| \frac{1}{m} \sum_{i=1}^{q} y_i y_i^* - \mathrm{I} \right\|_{2\to 2} \leqslant \frac{1}{2}$$

Application of Kadison-Singer problem

Lemma (N. Nagel, M. Schäfer and T. Ullrich)

Let $y_1, \dots, y_q \in \mathbb{C}^p$ such that $\|y_i\|_2^2 \leqslant k_1 \frac{p}{q}$ and

$$k_2 \, \mathrm{I} \leqslant \sum_{i=1}^{q} y_i y_i^* \leqslant k_3 \, \mathrm{I}$$

Then there is a $J \subset \{1, \dots, q\}$ such that $|J \cap \{1, \dots, n\}| \leqslant c_1 n \frac{p}{q}$ and

$$c_2 \frac{p}{q} \mathrm{I} \leqslant \sum_{i \in J} y_i y_i^* \leqslant c_3 \frac{p}{q} \mathrm{I}$$

Idea : Each time we split the sum in two, keep the partition class that has the fewest elements among $\{1, \dots, m\}$

Proof of the theorem

Denote $J' = J \cap \{1, \dots, m\}$

$$L = (b_k(x_i))_{i \in J', k < n} \quad \text{and} \quad \Phi = (\alpha a_k b_k(x_i))_{i \in J', k \geqslant n}$$

Then $|J'| \leqslant c_1 \frac{p}{q} \leqslant C_1 n$

$$L^* L \geqslant c_2 \frac{p}{q} \mathrm{I} \geqslant C_2 n \mathrm{I} \quad \text{and} \quad \Phi^* \Phi \leqslant c_3 \frac{p}{q} \mathrm{I} \leqslant C_3 n \mathrm{I}$$

After classical computations

$$\|f - Af\|_{L^2}^2 \leqslant a_n^2 + \|(L^* L)^{-1} L^*\|_{2 \to 2}^2 \, \|\Phi^* \Phi\|_{2 \to 2} \, \|f - Pf\|_{L^2}^2$$
$$\leqslant \left(1 + \frac{C_3}{C_2}\right) \alpha^{-2} \leqslant C \sum_{k \geqslant \lfloor n/2 \rfloor} a_k^2$$

and A uses $|J'| \leqslant C_1 n$ points, which concludes □

Thank you for your attention !