Implications of the Kadison Singer solution to the recovery of functions
– Optimal subsampling of random information–

Tino Ullrich

Faculty of Mathematics
Chemnitz University of Technology

Workshop: Sampling Recovery and Related Problems
Chemnitz/Moscow, May 2021

Joint work with...

Nicolas Nagel
Martin Schäfer

from the **Applied Analysis Group** at TU Chemnitz

Introduction

Recovery of functions in finite-dimensional spaces

- $V_m = \mathrm{span}\{\eta_1(\cdot), ..., \eta_m(\cdot)\} \subset L_2(D, \varrho)$, $f \in V_m$
- $(\eta_k(\cdot))_{k=1}^m$ ONS with respect to the measure ϱ
- **Goal:** Stable and exact recovery of $f \in V_m$ from given samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^T$
- Linear system

$$\begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}^1) \\ \vdots \\ f(\mathbf{x}^n) \end{pmatrix}$$

- How to choose the nodes $\mathbf{x}^1, ..., \mathbf{x}^n$ and the oversampling $n > m$ such that the system matrix is **well conditioned** ?
- **Weighted least squares**, change of measure, importance sampling, ...
- **Useful:** Bounded orthonormal systems (BOS) like trigonometric monomials, Chebychev polynomials, etc.

Recovery of functions in finite-dimensional spaces

- $V_m = \mathrm{span}\{\eta_1(\cdot), ..., \eta_m(\cdot)\} \subset L_2(D, \varrho)$, $f \in V_m$
- $(\eta_k(\cdot))_{k=1}^m$ ONS with respect to the measure ϱ
- **Goal:** Stable and exact recovery of $f \in V_m$ from given samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^T$
- Linear system

$$\begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}^1) \\ \vdots \\ f(\mathbf{x}^n) \end{pmatrix}$$

- How to choose the nodes $\mathbf{x}^1, ..., \mathbf{x}^n$ and the oversampling $n > m$ such that the system matrix is **well conditioned** ?
- **Weighted least squares**, change of measure, importance sampling, ...
- **Useful:** Bounded orthonormal systems (BOS) like trigonometric monomials, Chebychev polynomials, etc.

Recovery of functions in finite-dimensional spaces

Problem studied by many authors in more or less specific situations

- Finite-dimensional spaces of multivariate **trigonomeric/algebraic polynomials: Doostan, Gröchenig, Iwen, Kämmerer, Kunis, Krahmer, Mhaskar, Nobile, Potts, Rauhut, Temlyakov, Tempone, Volkmer, Ward, ...**

- General situation: **Cohen, Davenport, Leviatan, Migliorati, Adcock, ...**

- Marcinkiewicz-Zygmund inequalities and sampling discretization in $L_p(D,\varrho)$: **Dai, Shadrin, Tikhonov, Temlyakov + Lab People,**

$$c_1\|f\|_p^p \le \frac{1}{n}\sum_{k=1}^n |f(\mathbf{x}^k)|^p \le c_2\|f\|_p^p$$

  in case $p=2$: the lines of the above matrix constitute a proper frame in $\mathbb{C}^m$

Recovery of multivariate functions from RKHS

- **Model**: Reproducing kernel Hilbert space $H(K) \hookrightarrow L_2(D, \varrho_D)$
- **Given:** Samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^\top$ of a $f \in H(K)$
- **Goal:** Recover the function f from samples on $\mathbf{X}$
- **Additional assumption:** The sampling nodes $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ should work for a class of functions **simultaneously**
- We aim for controlling the **worst-case error** for a sampling recovery operator $S_{\mathbf{X}} : H(K) \to L_2$

$$\sup_{\|f\|_{H(K)} \leq 1} \|f - S_{\mathbf{X}} f\|_{L_2(D,\varrho)}$$

- **Information based complexity:** How well can we perform compared to general linear samples?
- **Sampling numbers** vs. **approximation numbers**

Recovery of multivariate functions from RKHS

- **Model**: Reproducing kernel Hilbert space $H(K) \hookrightarrow L_2(D, \varrho_D)$
- **Given:** Samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^\top$ of a $f \in H(K)$
- **Goal:** Recover the function f from samples on $\mathbf{X}$
- **Additional assumption:** The sampling nodes $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ should work for a class of functions **simultaneously**
- We aim for controlling the **worst-case error** for a sampling recovery operator $S_\mathbf{X} : H(K) \to L_2$

$$\sup_{\|f\|_{H(K)} \leq 1} \|f - S_\mathbf{X} f\|_{L_2(D,\varrho)}$$

- **Information based complexity:** How well can we perform compared to general linear samples?
- **Sampling numbers** vs. **approximation numbers**

Recovery of multivariate functions from RKHS

- **Model**: Reproducing kernel Hilbert space $H(K) \hookrightarrow L_2(D, \varrho_D)$
- **Given:** Samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^\top$ of a $f \in H(K)$
- **Goal:** Recover the function f from samples on $\mathbf{X}$
- **Additional assumption:** The sampling nodes $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ should work for a class of functions **simultaneously**
- We aim for controlling the **worst-case error** for a sampling recovery operator $S_\mathbf{X} : H(K) \to L_2$

$$\sup_{\|f\|_{H(K)} \leq 1} \|f - S_\mathbf{X} f\|_{L_2(D,\varrho)}$$

- **Information based complexity:** How well can we perform compared to general linear samples?
- **Sampling numbers** vs. **approximation numbers**

Recovery of multivariate functions from RKHS

- **Model**: Reproducing kernel Hilbert space $H(K) \hookrightarrow L_2(D, \varrho_D)$
- **Given:** Samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^\top$ of a $f \in H(K)$
- **Goal:** Recover the function f from samples on $\mathbf{X}$
- **Additional assumption:** The sampling nodes $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ should work for a class of functions **simultaneously**
- We aim for controlling the **worst-case error** for a sampling recovery operator $S_{\mathbf{X}} : H(K) \to L_2$

$$\sup_{\|f\|_{H(K)} \le 1} \|f - S_{\mathbf{X}} f\|_{L_2(D,\varrho)}$$

- **Information based complexity:** How well can we perform compared to general linear samples?
- **Sampling numbers** vs. **approximation numbers**

Weaver subsampling

It is known:

- General situation: $n = \mathcal{O}(m \log m)$ **random** samples
- In case of trigonometric polynomials $n = O(m^2)$ deterministic rank-1 lattice points

Reduce sampling budget to $\mathcal{O}(m)$

- Using the celebrated solution of Kadison-Singer via **Weaver's** conjecture (**2004**, Discr. Math.)
- **Nitzan, Olevskii, Ulanovskii, 2016:** Exponential frames on unbounded sets, Proc. Amer. Math. Soc.
- Solution of Kadison Singer Problem:
  Marcus, Spielman, Srivastava, 2015: Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. Ann. of Math.
- Forerunner to Kadison Singer solution
  Batson, Spielman, Srivastava, 2014: Twice Ramanujan Sparsifiers, SIAM Review

Weaver subsampling

It is known:

- General situation: $n = \mathcal{O}(m \log m)$ **random** samples
- In case of trigonometric polynomials $n = O(m^2)$ deterministic rank-1 lattice points

Reduce sampling budget to $\mathcal{O}(m)$

- Using the celebrated solution of Kadison-Singer via **Weaver's** conjecture (**2004**, Discr. Math.)
- **Nitzan, Olevskii, Ulanovskii, 2016:** Exponential frames on unbounded sets, Proc. Amer. Math. Soc.
- Solution of Kadison Singer Problem:
  Marcus, Spielman, Srivastava, 2015: Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem. Ann. of Math.
- Forerunner to Kadison Singer solution
  Batson, Spielman, Srivastava, 2014: Twice Ramanujan Sparsifiers, SIAM Review

Outline

Kadison Singer and Finite Frames

Frames and Riesz sequences

- Let $\mathcal{H}$ be a Hilbert space.
- A **Bessel sequence** is a sequence $(f_i)_{i\in I}$ in $\mathcal{H}$, such that there is a constant $C > 0$ with
$$\sum_{i\in I} |\langle f, f_i\rangle|^2 \le C\|f\|^2$$
for all $f \in \mathcal{H}$.
- A **Frame** is a sequence $(f_i)_{i\in I}$ in $\mathcal{H}$, such that there are constants $C, c > 0$ with
$$c\|f\|^2 \le \sum_{i\in I} |\langle f, f_i\rangle|^2 \le C\|f\|^2$$
for all $f \in \mathcal{H}$.
- A **Riesz sequence** is a sequence $(f_i)_{i\in I}$ in $\mathcal{H}$, for which there are constants $c, C > 0$, such that
$$c\sum_{i\in I} |a_i|^2 \le \left\|\sum_{i\in I} a_i f_i\right\|^2 \le C\sum_{i\in I} |a_i|^2$$

The Kadison Singer problem; Framework

- Motivated from Quantum Mechanics **Dirac 1947**
- $\mathcal{H} = \ell_2(\mathbb{N})$ Hilbert sequence space over $\mathbb{C}$.
- $\mathfrak{B} = \mathcal{L}(\mathcal{H})$ the space of bounded, linear operators $\mathcal{H} \to \mathcal{H}$.
- $\mathfrak{D} \subseteq \mathfrak{B}$ the space of diagonal operators (which forms a closed, unital C*-subalgebra).
- A **state** is a continuous, linear functional $\varphi : \mathfrak{D} \to \mathbb{C}$, such that
  (i) $\varphi(I) = 1$ (normalization);
  (ii) $\varphi(P) \geq 0$ for all positive operators $P \in \mathfrak{D}$ (positivity).
- Set of all states $\mathcal{S} \subseteq \mathfrak{D}'$ in the dual space of $\mathfrak{D}$ is convex, hence $\mathcal{S} \subseteq \mathfrak{D}'$ is the convex hull of its extreme points.
- Extreme points are called **pure states**, i.e. these are states, that cannot be written as a proper convex combination of at least two other states.

Extension of pure states

- Hahn-Banach $\Longrightarrow$ state φ on $\mathfrak{D}$ can be extended to $\mathfrak{B}$.
- **Kadison-Singer Problem 1959**: Is the extension of a pure state unique?

Theorem (Marcus, Spielman, Srivastava; 2015; Weaver 2004; Akemann, Anderson 1991)

Answer: YES

- It suffices to show: An extension of a pure state is zero, when evaluated on selfadjoint operators with zero-diagonal. It turned out that the following would be helpful.

Theorem (Anderson's Paving Conjecture 1991)

*For every $\varepsilon > 0$ there is an $r \in \mathbb{N}$ with the following property:
For every selfadjoint $T \in \mathbb{C}^{n\times n}$ with zero-diagonal, there are coordinate projections $P_1, ..., P_r$, such that $\sum_{i=1}^r P_i = I$ and $\|P_i T P_i\| \le \varepsilon \|T\|$ for all $i = 1, ..., r$.*

- Note, that the projections themself depend ond T, while the number of projections depends only on ε. One even has the bound $r \le 136/\varepsilon^4$.
- This can be pulled up to infinite matrices, from which one can deduce Kadison-Singer.
- Several equivalent formulations:

Kadison Singer $\Longleftrightarrow$ Anderson Paving $\Longleftrightarrow$ Weaver KS_2-conjecture $\Longleftrightarrow$ Feichtinger conjecture

What exactly did MSS prove?

Theorem (Marcus, Spielman, Srivastava '15)

Let $\varepsilon > 0$ and $\mathbf{v}_1, ..., \mathbf{v}_m \in \mathbb{C}^n$ be independent random vectors with finite support such that

$$\sum_{i=1}^{m} \mathbb{E}\mathbf{v}_i\mathbf{v}_i^* = \mathbf{I}$$

and $\mathbb{E}\|\mathbf{v}^i\|^2 \leq \varepsilon$ for all i, then $\mathbb{P}\Big(\Big\|\sum_{i=1}^{m}\mathbf{v}_i\mathbf{v}_i^*\Big\|_{2\to 2} \leq (1+\sqrt{\varepsilon})^2\Big) > 0$.

Let $r \in \mathbb{N}$ and let $\mathbf{u}_1, ..., \mathbf{u}_m \in \mathbb{C}^n$ be vectors fulfilling $\|\mathbf{u}_i\|_2^2 \leq \delta$ for all i and

$$\sum_{i=1}^{m} \mathbf{u}_i\mathbf{u}_i^* = \mathbf{I}.$$

Then there exits a partition $\{S_1, ..., S_r\}$ of $[m]$ such that

$$\Big\|\sum_{i\in S_j}\mathbf{u}_i\mathbf{u}_i^*\Big\|_{2\to 2} \leq \Big(1/\sqrt{r}+\sqrt{\delta}\Big)^2.$$

What exactly did MSS prove?

Theorem (Marcus, Spielman, Srivastava '15)

Let $\varepsilon > 0$ *and* $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{C}^n$ *be independent random vectors with finite support such that*

$$\sum_{i=1}^{m} \mathbb{E}\mathbf{v}_i\mathbf{v}_i^* = \mathbf{I}$$

and $\mathbb{E}\|\mathbf{v}^i\|^2 \leq \varepsilon$ *for all* i*, then* $\mathbb{P}\left(\left\|\sum_{i=1}^{m} \mathbf{v}_i\mathbf{v}_i^*\right\|_{2\to 2} \leq (1+\sqrt{\varepsilon})^2\right) > 0.$

Let $r \in \mathbb{N}$ and let $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{C}^n$ be vectors fulfilling $\|\mathbf{u}_i\|_2^2 \leq \delta$ for all i and

$$\sum_{i=1}^{m} \mathbf{u}_i\mathbf{u}_i^* = \mathbf{I}.$$

Then there exits a partition $\{S_1, \ldots, S_r\}$ of $[m]$ such that
$\left\|\sum_{i\in S_j} \mathbf{u}_i\mathbf{u}_i^*\right\|_{2\to 2} \leq \left(1/\sqrt{r}+\sqrt{\delta}\right)^2.$

Theorem (Weaver's KS_2-conjecture, Marcus Spielman Srivastava 2015)

There are universal constants $\eta \geq 2$ and $\theta > 0$, such that the following holds: Let $\mathbf{v}_1, ..., \mathbf{v}_m \in \mathbb{C}^n$ be unit vectors (i.e. $\|\mathbf{v}_i\| = 1$ in the euclidean norm for all $i = 1, ..., m$) and suppose

$$\sum_{i=1}^{m} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 = \eta \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ (η-tight frame). Then there is a partition $[m] = S_1 \dot{\cup} S_2$, such that

$$\sum_{i \in S_j} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 \leq (\eta - \theta) \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ and $j = 1, 2$ (frames with frame bounds θ and $\eta - \theta$).

Note that θ is only determined by η. An "isotropic collection" of vectors can be decomposed into two sets which are "approximately isotropic".

Explicit and quantitative version

Theorem (Nitzan, Olevskii, Ulanovskii 2016)

Let $\varepsilon > 0$ *and* $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{C}^n$ *with* $\|\mathbf{v}_i\|^2 \leq \varepsilon$ *and suppose*

$$\sum_{i=1}^{m} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 = \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ *(*η*-tight frame). Then there is a partition* $[m] = S_1 \dot\cup S_2$*, such that*

$$\sum_{i \in S_j} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 \leq \frac{(1+\sqrt{2\varepsilon})^2}{2} \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ *and* $j = 1, 2$*. In the case* $\varepsilon < 1$ *we therefore have*

$$\sum_{i \in S_j} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 \leq \frac{1+5\sqrt{\varepsilon}}{2} \|\mathbf{w}\|^2$$

Explicit and quantitative version

Theorem (Nitzan, Olevskii, Ulanovskii 2016)

Let $\varepsilon > 0$ *and* $\mathbf{v}_1, ..., \mathbf{v}_m \in \mathbb{C}^n$ *with* $\|\mathbf{v}_i\|^2 \leq \varepsilon$ *and suppose*

$$\sum_{i=1}^{m} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 = \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ *(η-tight frame). Then there is a partition* $[m] = S_1 \dot{\cup} S_2$*, such that*

$$\sum_{i \in S_j} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 \leq \frac{(1 + \sqrt{2\varepsilon})^2}{2} \|\mathbf{w}\|^2$$

for all $\mathbf{w} \in \mathbb{C}^n$ *and* $j = 1, 2$*. In the case* $\varepsilon < 1$ *we therefore have*

$$\sum_{i \in S_j} |\langle \mathbf{w}, \mathbf{v}_i \rangle|^2 \leq \frac{1 + 5\sqrt{\varepsilon}}{2} \|\mathbf{w}\|^2$$

Back to the recovery problem

- **Goal:** Stable and exact recovery of $f \in V_m$ from given samples $\mathbf{f} = (f(\mathbf{x}^1), ..., f(\mathbf{x}^n))^T$
- Linear system

$$\begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}^1) \\ \vdots \\ f(\mathbf{x}^n) \end{pmatrix}$$

- Goal: Reduce the sampling budget to $\mathcal{O}(m)$
- Random points and "change of measure" give $n = \mathcal{O}(m \log m)$
- **Example:** trigonometric polynomials with frequencies in index set $I \subset \mathbb{R}^d$

$$V_I = \text{span}\Big\{ \exp(i\mathbf{k} \cdot \mathbf{x}) \ : \ \mathbf{k} \in I \Big\} \quad , \quad |I| = m$$

$\mathcal{O}(m \log m)$ random samples or $n = \mathcal{O}(m^2)$ samples from rank-1 lattice...

Apply MSS inductively

Theorem (Nitzan, Olevskii, Ulanovskii 2016; Limonova, Temlyakov '20; Nagel, Schäfer, T. Ullrich' 20)

Let $k_1, k_2, k_3 > 0$ *and* $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{C}^m$ *with* $\|\mathbf{u}_i\|_2^2 \le k_1 \frac{m}{n}$ *for all* $i = 1, \dots, n$ *and*

$$k_2 \|\mathbf{w}\|_2^2 \le \sum_{i=1}^{n} |\langle \mathbf{w}, \mathbf{u}_i \rangle|^2 \le k_3 \|\mathbf{w}\|_2^2 \quad , \quad \mathbf{w} \in \mathbb{C}^m .$$

There always exists $J \subset \{1, \dots, n\}$ *with* $|J| \le c_1 \cdot m$ *such that*

$$c_2 \cdot \frac{m}{n} \|\mathbf{w}\|_2^2 \le \sum_{j \in J} |\langle \mathbf{w}, \mathbf{u}_j \rangle|^2 \le c_3 \cdot \frac{m}{n} \|\mathbf{w}\|_2^2 \, , \ \mathbf{w} \in \mathbb{C}^m ,$$

where $\quad c_1 = 1642 \frac{k_1}{k_2} \, , \quad c_2 = \min\{k_2, (2+\sqrt{2})^2 k_1\} \, , \quad c_3 = 1642 \frac{k_1 k_3}{k_2} \, .$

In its original NOU: $\|\mathbf{u}_i\|_2^2 = n/m$, **tight frame** and no explicit constants

Application to the recovery problem

- Denote the row vectors of $\mathbf{L}_m$ by $\mathbf{u}_1, \ldots, \mathbf{u}_n$:

$$\begin{pmatrix} ---\mathbf{u}_1 --- \\ \vdots \\ ---\mathbf{u}_n --- \end{pmatrix} := \mathbf{L}_m := \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}.$$

- $\left\{\frac{\mathbf{u}_1}{\sqrt{n}}, \ldots, \frac{\mathbf{u}_n}{\sqrt{n}}\right\}$ constitutes a frame in $\mathbb{C}^m$:

$$\frac{1}{2}\|\mathbf{w}\|_2^2 \leq \underbrace{\frac{1}{n}\left\|\mathbf{L}_m \mathbf{w}\right\|_2^2}_{\displaystyle \| \atop \displaystyle \sum_{i=1}^{n} \left|\left\langle \mathbf{w}, \frac{\mathbf{u}_i}{\sqrt{n}}\right\rangle\right|^2} \leq \frac{3}{2}\|\mathbf{w}\|_2^2$$

- Further, for all $i \in \{1, \ldots, n\}$: $\left\|\frac{\widetilde{\mathbf{u}}_i}{\sqrt{n}}\right\|_2^2 = \frac{m}{n}$ (after change of measure)

- We can apply Weaver's theorem with $k_1 = 2$, $k_2 = \frac{1}{2}$, $k_3 = \frac{1}{2}$.
- It yields a **subset** $J \subset \{1, \ldots, n\}$ and corresponding sampling points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ with $|J| \in \mathcal{O}(m)$ such that

$$c\|\mathbf{w}\|_2^2 \leq \frac{1}{m}\left\|\mathbf{L}_{\mathbf{J},m}\mathbf{w}\right\|_2^2 \leq C\|\mathbf{w}\|_2^2$$

for the following **submatrix** of $\mathbf{L}_m$:

$$\mathbf{L}_{\mathbf{J},m} := \begin{pmatrix} \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^j) & \eta_2(\mathbf{x}^j) & \cdots & \eta_m(\mathbf{x}^j) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{j \in J} = \begin{pmatrix} \vdots \\ ---\mathbf{u}_j--- \\ \vdots \end{pmatrix}_{j \in J} .$$

Three-step procedure

- **Change of measure $\Longrightarrow$ Random draw $\Longrightarrow$ Weaver subsampling**
- *1. Step:* **Change measure**, use density: $\varrho_m(\mathbf{x}) := \frac{1}{m}\sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$.

  Consequence: new system of orthonormal functions $\widetilde{\eta}_k(\cdot) := \eta_k(\cdot)/\sqrt{\varrho_m(\cdot)}$ with $\widetilde{N}(m) := \sup_{\mathbf{x}} \sum_{k=1}^{m} |\widetilde{\eta}_k(\mathbf{x})|^2 = m$
- *2. Step:* **Random draw** of $n = O(m \log m)$ points with respect to measure $\varrho_m(\mathbf{x})d\varrho_D(\mathbf{x})$. This constitutes a good **initial frame** with **bounded rows**
- *3. Step:* **Weaver subsampling** of the weighted matrix $\mathbf{D}_{\varrho_m} \cdot \mathbf{L}_m$ with

$$\mathbf{D}_{\varrho_m} = \mathrm{diag}(1/\sqrt{\varrho_m(\mathbf{x}^1)}, ..., 1/\sqrt{\varrho_m(\mathbf{x}^n)})$$

  Instead of solving

$$\mathbf{L}_m \cdot \mathbf{c} \approx \mathbf{f}$$

  we solve

$$\big(\mathbf{D}_{\varrho_m} \cdot \mathbf{L}_m\big) \cdot \tilde{\mathbf{c}} \approx \mathbf{D}_{\varrho_m} \cdot \mathbf{f}$$

Model Setting: RKHS

Reproducing kernel Hilbert space

- Problem setting by **Wasilkowski, Woźniakowski**, Journ. FoCM, 2001
- Domain $D \subset \mathbb{R}^d$ equipped with measure ϱ
- $K(\mathbf{x}, \mathbf{y})$ Hermitian positive-definite kernel on $D \times D$

$$f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_{H(K)} \quad (f \in H(K), \mathbf{x} \in D)$$

- Finite trace condition

$$\int_D K(\mathbf{x}, \mathbf{x}) \mathrm{d}\varrho < \infty.$$

- Hilbert-Schmidt embedding

$$\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D, \varrho)$$

- $\mathrm{Id}_{K,\varrho}$ is compact and its **singular values** $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are square-summable

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty$$

(Condition "necessary" for this setting: **Erich's lecture**)

Reproducing kernel Hilbert space

- Problem setting by **Wasilkowski, Woźniakowski**, Journ. FoCM, 2001
- Domain $D \subset \mathbb{R}^d$ equipped with measure ϱ
- $K(\mathbf{x}, \mathbf{y})$ Hermitian positive-definite kernel on $D \times D$

$$f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_{H(K)} \quad (f \in H(K),\, \mathbf{x} \in D)$$

- Finite trace condition

$$\int_D K(\mathbf{x}, \mathbf{x}) \mathrm{d}\varrho < \infty.$$

- Hilbert-Schmidt embedding

$$\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D, \varrho)$$

- $\mathrm{Id}_{K,\varrho}$ is compact and its **singular values** $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are square-summable

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty$$

(Condition "necessary" for this setting: **Erich's lecture**)

Reproducing kernel Hilbert space

- Problem setting by **Wasilkowski, Woźniakowski**, Journ. FoCM, 2001
- Domain $D \subset \mathbb{R}^d$ equipped with measure ϱ
- $K(\mathbf{x}, \mathbf{y})$ Hermitian positive-definite kernel on $D \times D$

$$f(\mathbf{x}) = \langle f, K(\cdot, \mathbf{x}) \rangle_{H(K)} \quad (f \in H(K),\, \mathbf{x} \in D)$$

- Finite trace condition

$$\int_D K(\mathbf{x}, \mathbf{x}) \mathrm{d}\varrho < \infty.$$

- Hilbert-Schmidt embedding

$$\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D, \varrho)$$

- $\mathrm{Id}_{K,\varrho}$ is compact and its **singular values** $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are square-summable

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty$$

(Condition "necessary" for this setting: **Erich's lecture**)

Reproducing kernel Hilbert space

- Problem setting by **Wasilkowski, Woźniakowski**, Journ. FoCM, 2001
- Domain $D \subset \mathbb{R}^d$ equipped with measure ϱ
- $K(\mathbf{x},\mathbf{y})$ Hermitian positive-definite kernel on $D \times D$

$$f(\mathbf{x}) = \langle f, K(\cdot,\mathbf{x})\rangle_{H(K)} \quad (f \in H(K),\, \mathbf{x} \in D)$$

- Finite trace condition

$$\int_D K(\mathbf{x},\mathbf{x})\mathrm{d}\varrho < \infty.$$

- Hilbert-Schmidt embedding

$$\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$$

- $\mathrm{Id}_{K,\varrho}$ is compact and its **singular values** $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are square-summable

$$\sum_{k=1}^{\infty} \sigma_k^2 < \infty$$

(Condition "necessary" for this setting: **Erich's lecture**)

Reproducing kernel Hilbert space

- $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$ has the following representation

$$\mathrm{Id}_{K,\varrho}(f) = \sum_{k=1}^{\infty} \sigma_k \langle f, e_k \rangle_{H(K)} \eta_k$$

- $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are the **singular values** of $\mathrm{Id}_{K,\varrho}$.
- $\{e_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are the **right** resp. **left singular functions**

$$\{e_k\}_{k\in\mathbb{N}} \text{ is an ONS in } H(K) : \quad \langle e_j, e_k \rangle_{H(K)} = \delta_{j,k}$$
$$\{\eta_k\}_{k\in\mathbb{N}} \text{ is an ONS in} L_2(D,\varrho) : \quad \langle \eta_j, \eta_k \rangle_{L_2(D,\varrho)} = \delta_{j,k}$$

- Since $\mathrm{Id}_{K,\varrho}$ is the identity on functions

$$e_k = \sigma_k \cdot \eta_k$$

Reproducing kernel Hilbert space

- $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$ has the following representation

$$\mathrm{Id}_{K,\varrho}(f) = \sum_{k=1}^{\infty} \sigma_k \langle f, e_k \rangle_{H(K)} \eta_k$$

- $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are the **singular values** of $\mathrm{Id}_{K,\varrho}$.
- $\{e_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are the **right** resp. **left singular functions**

$\{e_k\}_{k\in\mathbb{N}}$ is an ONS in $H(K)$: $\langle e_j, e_k \rangle_{H(K)} = \delta_{j,k}$
$\{\eta_k\}_{k\in\mathbb{N}}$ is an ONS in $L_2(D,\varrho)$: $\langle \eta_j, \eta_k \rangle_{L_2(D,\varrho)} = \delta_{j,k}$

- Since $\mathrm{Id}_{K,\varrho}$ is the identity on functions

$$e_k = \sigma_k \cdot \eta_k$$

Reproducing kernel Hilbert space

- $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$ has the following representation

$$\mathrm{Id}_{K,\varrho}(f) = \sum_{k=1}^{\infty} \sigma_k \langle f, e_k \rangle_{H(K)} \eta_k$$

- $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are the **singular values** of $\mathrm{Id}_{K,\varrho}$.
- $\{e_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are the **right** resp. **left singular functions**

$\{e_k\}_{k\in\mathbb{N}}$ is an ONS in $H(K)$: $\quad \langle e_j, e_k \rangle_{H(K)} = \delta_{j,k}$

$\{\eta_k\}_{k\in\mathbb{N}}$ is an ONS in$L_2(D,\varrho)$: $\quad \langle \eta_j, \eta_k \rangle_{L_2(D,\varrho)} = \delta_{j,k}$

- Since $\mathrm{Id}_{K,\varrho}$ is the identity on functions

$$e_k = \sigma_k \cdot \eta_k$$

Reproducing kernel Hilbert space

- $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$ has the following representation

$$\mathrm{Id}_{K,\varrho}(f) = \sum_{k=1}^{\infty} \sigma_k \langle f, e_k \rangle_{H(K)} \eta_k$$

- $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq 0$ are the **singular values** of $\mathrm{Id}_{K,\varrho}$.
- $\{e_k\}_{k\in\mathbb{N}}$ and $\{\eta_k\}_{k\in\mathbb{N}}$ are the **right** resp. **left singular functions**

$\{e_k\}_{k\in\mathbb{N}}$ is an ONS in $H(K)$: $\quad \langle e_j, e_k \rangle_{H(K)} = \delta_{j,k}$

$\{\eta_k\}_{k\in\mathbb{N}}$ is an ONS in $L_2(D,\varrho)$: $\quad \langle \eta_j, \eta_k \rangle_{L_2(D,\varrho)} = \delta_{j,k}$

- Since $\mathrm{Id}_{K,\varrho}$ is the identity on functions

$$e_k = \sigma_k \cdot \eta_k$$

How to approximate?

Algorithm 1 Weighted least squares approximation

Input:	$\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^n) \in D^n$	matrix of distinct sampling nodes,
	$\mathbf{f} = (f(\mathbf{x}^1), \dots, f(\mathbf{x}^n))^\top$	samples of f evaluated at the nodes,
	$m \in \mathbb{N}$	$m \le n$

Compute weighted samples

$$\boldsymbol{g} := (g_j)_{j=1}^n \text{ with } g_j := \begin{cases} 0, & \varrho_m(\mathbf{x}^j) = 0, \\ f(\mathbf{x}^j)/\sqrt{\varrho_m(\mathbf{x}^j)}, & \varrho_m(\mathbf{x}^j) \neq 0\,. \end{cases}$$

Least squares matrix

$$l_{j,\ell} := \begin{cases} 0, & \varrho_m(\mathbf{x}^j) = 0, \\ \eta_\ell(\mathbf{x}^j)/\sqrt{\varrho_m(\mathbf{x}^j)}, & \varrho_m(\mathbf{x}^j) \neq 0, \end{cases} \tag{1}$$

Output coefficients:

$$(\tilde{c}_1, \dots, \tilde{c}_m)^\top := (\widetilde{\mathbf{L}}_{n,m}^* \widetilde{\mathbf{L}}_{n,m})^{-1}\, \widetilde{\mathbf{L}}_{n,m}^* \cdot \mathbf{g}.$$

A concrete example

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$

$$\langle f, g\rangle_{H^r_{\text{mix}}} := \sum_{\mathbf{j}\in\{0,r\}^d} \langle D^{(\mathbf{j})} f, D^{(\mathbf{j})} g\rangle_{L_2(\mathbb{T}^d)} \,.$$

- $D = \mathbb{T}^d \mathrel{\widehat{=}} [0,1]^d$, $\mathrm{d}\varrho(\mathbf{x}) = \mathrm{d}\mathbf{x}$
- $r > \frac{1}{2}$

$$w_s(k) = (1 + (2\pi|k|)^{2r})^{1/2} \quad , \quad k \in \mathbb{Z}$$

$$K^1_r(x,y) := \sum_{k\in\mathbb{Z}} \frac{\exp(2\pi \mathrm{i} k(y-x))}{w_r(k)^2} \quad , \quad x, y \in \mathbb{T}\,,$$

$$K^d_r(\mathbf{x},\mathbf{y}) := K^1_r(x_1,y_1) \otimes \cdots \otimes K^1_r(x_d,y_d) \quad , \quad \mathbf{x},\mathbf{y} \in \mathbb{T}^d\,,$$

- Singular numbers $\sigma_n = (1/w_r(\mathbf{k}_n))_n$ (non-increasing rearrangement)
- $e_n(\mathbf{x}) = \sigma_n \exp(2\pi i \mathbf{k}_n \cdot)$, $\eta_n(\mathbf{x}) = \exp(2\pi i \mathbf{k}_n \cdot)$

A concrete example

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$

$$\langle f, g\rangle_{H^r_{\text{mix}}} := \sum_{\mathbf{j}\in\{0,r\}^d} \langle D^{(\mathbf{j})} f, D^{(\mathbf{j})} g\rangle_{L_2(\mathbb{T}^d)}\,.$$

- $D = \mathbb{T}^d \mathrel{\widehat{=}} [0,1]^d$, $\mathrm{d}\varrho(\mathbf{x}) = \mathrm{d}\mathbf{x}$
- $r > \frac{1}{2}$

$$w_s(k) = (1 + (2\pi|k|)^{2r})^{1/2} \quad , \quad k \in \mathbb{Z}$$

$$K^1_r(x,y) := \sum_{k\in\mathbb{Z}} \frac{\exp(2\pi \mathrm{i} k(y-x))}{w_r(k)^2} \quad , \quad x,y \in \mathbb{T}\,,$$

$$K^d_r(\mathbf{x},\mathbf{y}) := K^1_r(x_1,y_1) \otimes \cdots \otimes K^1_r(x_d,y_d) \quad , \quad \mathbf{x},\mathbf{y} \in \mathbb{T}^d\,,$$

- Singular numbers $\sigma_n = (1/w_r(\mathbf{k}_n))_n$ (non-increasing rearrangement)
- $e_n(\mathbf{x}) = \sigma_n \exp(2\pi i \mathbf{k}_n\cdot)$, $\eta_n(\mathbf{x}) = \exp(2\pi i \mathbf{k}_n\cdot)$

A concrete example

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$

$$\langle f, g\rangle_{H^r_{\text{mix}}} := \sum_{\mathbf{j}\in\{0,r\}^d} \langle D^{(\mathbf{j})} f, D^{(\mathbf{j})} g\rangle_{L_2(\mathbb{T}^d)} \,.$$

- $D = \mathbb{T}^d \mathrel{\widehat{=}} [0,1]^d$, $\mathrm{d}\varrho(\mathbf{x}) = \mathrm{d}\mathbf{x}$
- $r > \frac{1}{2}$

$$w_s(k) = (1 + (2\pi|k|)^{2r})^{1/2} \quad , \quad k \in \mathbb{Z}$$

$$K^1_r(x,y) := \sum_{k\in\mathbb{Z}} \frac{\exp(2\pi \mathrm{i} k(y-x))}{w_r(k)^2} \quad , \quad x,y \in \mathbb{T}\,,$$

$$K^d_r(\mathbf{x},\mathbf{y}) := K^1_r(x_1,y_1) \otimes \cdots \otimes K^1_r(x_d,y_d) \quad , \quad \mathbf{x},\mathbf{y} \in \mathbb{T}^d\,,$$

- Singular numbers $\sigma_n = (1/w_r(\mathbf{k}_n))_n$ (non-increasing rearrangement)
- $e_n(\mathbf{x}) = \sigma_n \exp(2\pi i \mathbf{k}_n \cdot)$, $\eta_n(\mathbf{x}) = \exp(2\pi i \mathbf{k}_n \cdot)$

A concrete example

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$

$$\langle f, g\rangle_{H^r_{\text{mix}}} := \sum_{\mathbf{j}\in\{0,r\}^d} \langle D^{(\mathbf{j})} f, D^{(\mathbf{j})} g\rangle_{L_2(\mathbb{T}^d)} \,.$$

- $D = \mathbb{T}^d \mathrel{\widehat{=}} [0,1]^d$, $\mathrm{d}\varrho(\mathbf{x}) = \mathrm{d}\mathbf{x}$
- $r > \frac{1}{2}$

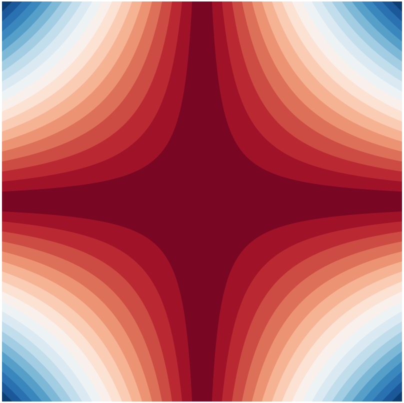
$$w_s(k) = (1 + (2\pi|k|)^{2r})^{1/2} \quad , \quad k \in \mathbb{Z}$$

$$K^1_r(x,y) := \sum_{k\in\mathbb{Z}} \frac{\exp(2\pi\mathrm{i}k(y-x))}{w_r(k)^2} \quad , \quad x, y \in \mathbb{T}\,,$$

$$K^d_r(\mathbf{x},\mathbf{y}) := K^1_r(x_1,y_1) \otimes \cdots \otimes K^1_r(x_d,y_d) \quad , \quad \mathbf{x},\mathbf{y} \in \mathbb{T}^d\,,$$

- Singular numbers $\sigma_n = (1/w_r(\mathbf{k}_n))_n$ (non-increasing rearrangement)
- $e_n(\mathbf{x}) = \sigma_n \exp(2\pi i \mathbf{k}_n \cdot)$, $\eta_n(\mathbf{x}) = \exp(2\pi i \mathbf{k}_n \cdot)$

The "smooth" hyperbolic cross



- Special case $p = q = 2$

$$\|f\|^2_{H^r_{mix}} := \sum_{k \in \mathbb{Z}} |\hat{f}(k)|^2 \prod_{i=1}^{d} (1+|k_i|^2)^r$$

- Hyperbolic cross projection

$$P_{\mathcal{H}_n} := \sum_{k \in \mathcal{H}_n} \hat{f}(k) e^{2\pi i k x}$$

- **Error:** $\|f - P_{\mathcal{H}_n} f\|_2 \le n^{-r}$
- **Cost:** $m := \sharp$ grid points in $\mathcal{H}_n$
- **Rate:** $m^{-r} (\log m)^{(d-1)r}$

Matrix formulation of the recovery error

- Error analysis:

$$\sup_{\|f\|_{H(K)}\le 1} \|f - S^m_{\mathbf{X}} f\|^2_{L_2} \le \sup \|f - P_m f\|^2_{L_2} + \sup \|S^m_{\mathbf{X}}(f - P_m f)\|^2_{L_2}$$

$$\le \sigma^2_{m+1} + \frac{\sigma_{\max}(\Phi_m)^2}{\sigma_{\min}(\mathbf{L}_m)^2}$$

- **Least squares matrix:**

$$\mathbf{L}_m := \mathbf{L}_{\mathbf{X},m} := \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Remainder:** Infinite matrix

$$\Phi_m := \Phi_{\mathbf{X},m} := \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

Matrix formulation of the recovery error

- Error analysis:

$$\sup_{\|f\|_{H(K)}\le 1} \|f - S^m_{\mathbf{X}} f\|^2_{L_2} \le \sup \|f - P_m f\|^2_{L_2} + \sup \|S^m_{\mathbf{X}}(f - P_m f)\|^2_{L_2}$$

$$\le \sigma^2_{m+1} + \frac{\sigma_{\max}(\Phi_m)^2}{\sigma_{\min}(\mathbf{L}_m)^2}$$

- **Least squares matrix:**

$$\mathbf{L}_m := \mathbf{L}_{\mathbf{X},m} := \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Remainder:** Infinite matrix

$$\Phi_m := \Phi_{\mathbf{X},m} := \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- We want to construct sampling nodes $\mathbf{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$ probabilistically such that

- $\ldots\ \sigma_{\min}(\mathbf{L}_m)$ is large for $\quad \mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$

- $\ldots\ \sigma_{\max}(\Phi_m)$ is small for $\quad \Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$

- Previous work on this subject, e.g.

Krieg, M. Ullrich '19
Cohen, Migliorati '16

Cohen, Davenport, Leviatan '13

- We want to construct sampling nodes $\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^n)$ probabilistically such that
- $\dots$ $\sigma_{\min}(\mathbf{L}_m)$ is large for $$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$
- $\dots$ $\sigma_{\max}(\Phi_m)$ is small for $$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$
- Previous work on this subject, e.g.

  Krieg, M. Ullrich '19
  Cohen, Migliorati '16

  Cohen, Davenport, Leviatan '13

- We want to construct sampling nodes $\mathbf{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$ probabilistically such that
- $\ldots\ \sigma_{\min}(\mathbf{L}_m)$ is large for $$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$
- $\ldots\ \sigma_{\max}(\Phi_m)$ is small for $$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$
- Previous work on this subject, e.g.

  Krieg, M. Ullrich '19
  Cohen, Migliorati '16

  Cohen, Davenport, Leviatan '13

- We want to construct sampling nodes $\mathbf{X} = (\mathbf{x}^1, \ldots, \mathbf{x}^n)$ probabilistically such that
- $\ldots\ \sigma_{\min}(\mathbf{L}_m)$ is large for $\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$
- $\ldots\ \sigma_{\max}(\Phi_m)$ is small for $\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$
- Previous work on this subject, e.g.

Krieg, M. Ullrich '19
Cohen, Migliorati '16

Cohen, Davenport, Leviatan '13

Random Matrices

- $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ is drawn i.i.d. at random.
- Random matrix

$$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded by "inverse Christoffel function"

$$N(m) := \sup_{\mathbf{x} \in D} \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$$

- Random matrix

$$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded

- $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ is drawn i.i.d. at random.
- Random matrix

$$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded by “inverse Christoffel function”

$$N(m) := \sup_{\mathbf{x} \in D} \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$$

- Random matrix

$$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded

- $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ is drawn i.i.d. at random.
- Random matrix

$$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded by "inverse Christoffel function"

$$N(m) := \sup_{\mathbf{x}\in D} \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$$

- Random matrix

$$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded

- $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ is drawn i.i.d. at random.
- Random matrix

$$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded by "inverse Christoffel function"

$$N(m) := \sup_{\mathbf{x} \in D} \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$$

- Random matrix

$$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded

- $\mathbf{X} = (\mathbf{x}^1, ..., \mathbf{x}^n)$ is drawn i.i.d. at random.
- Random matrix

$$\mathbf{L}_m = \begin{pmatrix} \eta_1(\mathbf{x}^1) & \eta_2(\mathbf{x}^1) & \cdots & \eta_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \eta_1(\mathbf{x}^n) & \eta_2(\mathbf{x}^n) & \cdots & \eta_m(\mathbf{x}^n) \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded by "inverse Christoffel function"

$$N(m) := \sup_{\mathbf{x} \in D} \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$$

- Random matrix

$$\Phi_m = \begin{pmatrix} e_{m+1}(\mathbf{x}^1) & e_{m+2}(\mathbf{x}^1) & \cdots \\ \vdots & \vdots & \\ e_{m+1}(\mathbf{x}^n) & e_{m+2}(\mathbf{x}^n) & \cdots \end{pmatrix}$$

- **Structure:** Rows i.i.d. and ℓ_2-bounded

Random matrices

- Take independent samples according to ϱ.
- Then

$$\mathbb{E}\Big(\frac{1}{n}\mathbf{L}_m^*\mathbf{L}_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{y}^k \otimes \mathbf{y}^k\Big) = \mathbf{I}_m$$

- Further

$$\mathbb{E}\Big(\frac{1}{n}\Phi_m^*\Phi_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}^k \otimes \mathbf{z}^k\Big) = \mathrm{diag}(\sigma_{m+1}^2, \sigma_{m+2}^2, \ldots) =: \mathbf{\Lambda}_m$$

- Recall

$$\sigma_{\min}^2(\mathbf{L}_m) = \lambda_{\min}(\mathbf{L}_m^*\mathbf{L}_m)$$
$$\sigma_{\max}^2(\Phi_m) = \lambda_{\max}(\Phi_m^*\Phi_m)$$

-

Random matrices

- Take independent samples according to ϱ.
- Then

$$\mathbb{E}\Big(\frac{1}{n}\mathbf{L}_m^*\mathbf{L}_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{y}^k \otimes \mathbf{y}^k\Big) = \mathbf{I}_m$$

- Further

$$\mathbb{E}\Big(\frac{1}{n}\Phi_m^*\Phi_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}^k \otimes \mathbf{z}^k\Big) = \operatorname{diag}(\sigma_{m+1}^2, \sigma_{m+2}^2, \ldots) =: \mathbf{\Lambda}_m$$

- Recall

$$\sigma_{\min}^2(\mathbf{L}_m) = \lambda_{\min}(\mathbf{L}_m^*\mathbf{L}_m)$$
$$\sigma_{\max}^2(\Phi_m) = \lambda_{\max}(\Phi_m^*\Phi_m)$$

Random matrices

- Take independent samples according to ϱ.
- Then

$$\mathbb{E}\Big(\frac{1}{n}\mathbf{L}_m^*\mathbf{L}_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{y}^k \otimes \mathbf{y}^k\Big) = \mathbf{I}_m$$

- Further

$$\mathbb{E}\Big(\frac{1}{n}\Phi_m^*\Phi_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}^k \otimes \mathbf{z}^k\Big) = \operatorname{diag}(\sigma_{m+1}^2, \sigma_{m+2}^2, \ldots) =: \mathbf{\Lambda}_m$$

- Recall

$$\sigma_{\min}^2(\mathbf{L}_m) = \lambda_{\min}(\mathbf{L}_m^*\mathbf{L}_m)$$
$$\sigma_{\max}^2(\Phi_m) = \lambda_{\max}(\Phi_m^*\Phi_m)$$

Random matrices

- Take independent samples according to ϱ.
- Then

$$\mathbb{E}\Big(\frac{1}{n}\mathbf{L}_m^*\mathbf{L}_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{y}^k \otimes \mathbf{y}^k\Big) = \mathbf{I}_m$$

- Further

$$\mathbb{E}\Big(\frac{1}{n}\Phi_m^*\Phi_m\Big) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}^k \otimes \mathbf{z}^k\Big) = \operatorname{diag}(\sigma_{m+1}^2, \sigma_{m+2}^2, \ldots) =: \mathbf{\Lambda}_m$$

- Recall

$$\sigma_{\min}^2(\mathbf{L}_m) = \lambda_{\min}(\mathbf{L}_m^*\mathbf{L}_m)$$
$$\sigma_{\max}^2(\Phi_m) = \lambda_{\max}(\Phi_m^*\Phi_m)$$

-

Matrix Chernoff (Tropp)

Theorem (Tropp 2010)

For a finite sequence $(\mathbf{A}_k)$ *of independent, self-adjoint, positive semi-definite random matrices with dimension* m *satisfying* $\lambda_{\max}(\mathbf{A}_k) \le R$ *almost surely it holds*

$$\mathbb{P}\Big(\lambda_{\min}\Big(\sum_{k=1}^{n} \mathbf{A}_k\Big) \le (1-t)\mu_{\min}\Big) \le m\Big(\frac{e^{-t}}{(1-t)^{1-t}}\Big)^{\frac{\mu_{\min}}{R}}$$

$$\mathbb{P}\Big(\lambda_{\max}\Big(\sum_{k=1}^{n} \mathbf{A}_k\Big) \ge (1+t)\mu_{\max}\Big) \le m\Big(\frac{e^{t}}{(1+t)^{1+t}}\Big)^{\frac{\mu_{\max}}{R}}$$

for $t \in [0,1]$ *where* $\mu_{\min} := \lambda_{\min}\left(\sum_{k=1}^{m} \mathbb{E}\mathbf{A}_k\right)$ *and* $\mu_{\max} := \lambda_{\max}\left(\sum_{k=1}^{m} \mathbb{E}\mathbf{A}_k\right)$.

Apply with: $\mu_{\min} = \mu_{\max} = n$, $t = 1/2$, $R = N(m) = \sum_{k=1}^{m} |\eta_k(\mathbf{x})|^2$

Infinite matrices

Proposition (Moeller, T. Ullrich 2020)

Let $\mathbf{y}^i, i = 1 \ldots n$, *be i.i.d random sequences from* ℓ_2. *Let further* $n \geq 3$, $M > 0$ *such that* $\|\mathbf{z}^i\|_2 \leq M$ *for all* $i = 1 \ldots n$ *almost surely and* $\mathbb{E}\mathbf{z}^i \otimes \mathbf{z}^i = \mathbf{\Lambda}$ *for* $i = 1, \ldots, n$. *Then*

$$\mathbb{P}\left(\left\|\frac{1}{n}\sum_{i=1}^{n} \mathbf{z}^i \otimes \mathbf{z}^i - \mathbf{\Lambda}\right\|_{2\to 2} \geq \max\left\{\|\mathbf{\Lambda}\|_{2\to 2}, 21r\frac{\log n}{n}M^2\right\}\right) \leq 2^{\frac{3}{4}} n^{1-r}$$

- Complements earlier results by **Rauhut, Pajor, Mendelson, Rudelson, Oliveira...**
- Focus here on **infinite** random matrices, but also valid for random vectors of fixed finite length
- Proof based on non-commutative Khintchine inequality (**Buchholz**)

Recovery with high probability

New sampling measure

- **Importance sampling**: Sample more densely, where samples are more relevant.

  $\rightsquigarrow$ Modify the $\mathrm{d}\varrho(\mathbf{x})$ on D with an appropriate density function.

  • First approach due to **Cohen, Migliorati '16** (weighted least squares, change of measure)

- **Krieg, M. Ullrich '19** suggested the density

$$\varrho_m(\mathbf{x}) := \frac{1}{2}\left(\frac{1}{m}\sum_{k=1}^{m}|\eta_k(\mathbf{x})|^2 + \frac{\sum_{k=m+1}^{\infty}|e_k(\mathbf{x})|^2}{\sum_{k=m+1}^{\infty}\sigma_k^2}\right)$$

- New sampling measure

$$\mathrm{d}\widetilde{\varrho}_m(\mathbf{x}) := \varrho_m(\mathbf{x}) \cdot \mathrm{d}\varrho(\mathbf{x})$$

New sampling measure

- **Importance sampling**: Sample more densely, where samples are more relevant.

  $\rightsquigarrow$ Modify the $\mathrm{d}\varrho(\mathbf{x})$ on D with an appropriate density function.

  • First approach due to **Cohen, Migliorati '16** (weighted least squares, change of measure)
- **Krieg, M. Ullrich '19** suggested the density

$$\varrho_m(\mathbf{x}) := \frac{1}{2}\left(\frac{1}{m}\sum_{k=1}^{m}|\eta_k(\mathbf{x})|^2 + \frac{\sum_{k=m+1}^{\infty}|e_k(\mathbf{x})|^2}{\sum_{k=m+1}^{\infty}\sigma_k^2}\right)$$

- New sampling measure

$$\mathrm{d}\widetilde{\varrho}_m(\mathbf{x}) := \varrho_m(\mathbf{x}) \cdot \mathrm{d}\varrho(\mathbf{x})$$

New sampling measure

- **Importance sampling**: Sample more densely, where samples are more relevant.

  $\rightsquigarrow$ Modify the $\mathrm{d}\varrho(\mathbf{x})$ on D with an appropriate density function.

  • First approach due to **Cohen, Migliorati '16** (weighted least squares, change of measure)
- **Krieg, M. Ullrich '19** suggested the density

$$\varrho_m(\mathbf{x}) := \frac{1}{2}\left(\frac{1}{m}\sum_{k=1}^{m}|\eta_k(\mathbf{x})|^2 + \frac{\sum_{k=m+1}^{\infty}|e_k(\mathbf{x})|^2}{\sum_{k=m+1}^{\infty}\sigma_k^2}\right)$$

- New sampling measure

$$\mathrm{d}\widetilde{\varrho}_m(\mathbf{x}) := \varrho_m(\mathbf{x}) \cdot \mathrm{d}\varrho(\mathbf{x})$$

New sampling measure

- **Importance sampling**: Sample more densely, where samples are more relevant.

  $\rightsquigarrow$ Modify the $\mathrm{d}\varrho(\mathbf{x})$ on D with an appropriate density function.

  • First approach due to **Cohen, Migliorati '16** (weighted least squares, change of measure)

- **Krieg, M. Ullrich '19** suggested the density

$$\varrho_m(\mathbf{x}) := \frac{1}{2}\left(\frac{1}{m}\sum_{k=1}^{m}|\eta_k(\mathbf{x})|^2 + \frac{\sum_{k=m+1}^{\infty}|e_k(\mathbf{x})|^2}{\sum_{k=m+1}^{\infty}\sigma_k^2}\right)$$

- New sampling measure

$$\mathrm{d}\widetilde{\varrho}_m(\mathbf{x}) := \varrho_m(\mathbf{x}) \cdot \mathrm{d}\varrho(\mathbf{x})$$

Weighted least squares

Trick: Instead of solving

$$\mathbf{L}_m \cdot \mathbf{c} \approx \mathbf{f}$$

we solve with $\mathbf{D}_{\varrho_m} = \mathrm{diag}(1/\sqrt{\varrho_m(\mathbf{x}^1)}, ..., 1/\sqrt{\varrho_m(\mathbf{x}^n)})$

$$\big(\mathbf{D}_{\varrho_m} \cdot \mathbf{L}_m\big) \cdot \tilde{\mathbf{c}} \approx \mathbf{D}_{\varrho_m} \cdot \mathbf{f}$$

and

$$\widetilde{N}(m) \sim m \quad , \quad \widetilde{T}(m) \sim \sum_{k=m+1}^{\infty} \sigma_k^2 \, .$$

$\Longrightarrow n \sim m \log m$! Similar behavior as for bounded orthonormal systems.

- **M. Ullrich '20:** high probability version based on **Oliveira's** spectral concentration result
- Direct consequence of our new infinite matrix concentration result above...

Theorem (Moeller, T. Ullrich 2020)

Let K be a positive definite kernel such that $H(K)$ is separable and $\int_D K(\mathbf{x},\mathbf{x})d\varrho(\mathbf{x}) < \infty$. Let $n \in \mathbb{N}$ and

$$m := \left\lfloor \frac{n}{14r\log n} \right\rfloor .$$

Then it holds

$$\mathbb{P}\Big(\sup_{\|f\|_{H(K)}\le 1} \|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho_D)} \le \frac{15}{m} \sum_{j=\lfloor m/2 \rfloor}^{\infty} \sigma_j^2\Big) \ge 1 - 3n^{1-r} ,$$

where $\mathbf{x}^i$, $i = 1, ..., n$, are sampled independently according to the measure $\varrho_m(\mathbf{x})d\varrho_D(\mathbf{x})$.

- **M. Ullrich '20:** high probability version based on **Oliveira's** spectral concentration result
- Direct consequence of our new infinite matrix concentration result above...

Theorem (Moeller, T. Ullrich 2020)

Let K be a positive definite kernel such that $H(K)$ is separable and $\int_D K(\mathbf{x},\mathbf{x})d\varrho(\mathbf{x}) < \infty$. Let $n \in \mathbb{N}$ and

$$m := \left\lfloor \frac{n}{14r\log n} \right\rfloor .$$

Then it holds

$$\mathbb{P}\Big(\sup_{\|f\|_{H(K)}\le 1} \|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho_D)} \le \frac{15}{m} \sum_{j=\lfloor m/2 \rfloor}^{\infty} \sigma_j^2 \Big) \ge 1 - 3n^{1-r} ,$$

where $\mathbf{x}^i$, $i = 1, ..., n$, are sampled independently according to the measure $\varrho_m(\mathbf{x})d\varrho_D(\mathbf{x})$.

Weaver Subsampling in RKHS

Apply Weaver subsampling to RKHS

- Random approach requires $\mathcal{O}(m\log(m))$ samples for a reasonable approximation ($\Longrightarrow$ least squares system matrix $\mathbf{L}_m$ with $m\log m$ rows
- We can "shrink" the matrix $\mathbf{L}_m$ to $\mathcal{O}(m)$ lines
  $\Longrightarrow$ **Weaver subsampling**, see also **Sparsification**, etc.
- Instead of operator $\widetilde{S}^m_{\mathbf{X}}$ we use the operator $\widetilde{S}^m_{\mathbf{J}}$, where $\mathbf{J}$ refers to the relevant samples
- **Consequence:** There exists a **node subset** of size $\mathcal{O}(m)$ such that

$$\sup_{\|f\|_{H(K)}\le 1} \|f - \widetilde{S}^m_{\mathbf{J}} f\|^2_{L_2(D,\varrho)} \le C\frac{\log(m)}{m}\sum_{k=\lfloor cm\rfloor}^{\infty}\sigma_k^2$$

  $\Longrightarrow$ We produce an additional $\sqrt{\log m}$
- The constants $C, c > 0$ are universal and determined explicitly!

Apply Weaver subsampling to RKHS

- Random approach requires $\mathcal{O}(m\log(m))$ samples for a reasonable approximation ($\Longrightarrow$ least squares system matrix $\mathbf{L}_m$ with $m\log m$ rows
- We can "shrink" the matrix $\mathbf{L}_m$ to $\mathcal{O}(m)$ lines
  $\Longrightarrow$ **Weaver subsampling**, see also **Sparsification**, etc.
- Instead of operator $\widetilde{S}^m_{\mathbf{X}}$ we use the operator $\widetilde{S}^m_{\mathbf{J}}$, where $\mathbf{J}$ refers to the relevant samples
- **Consequence:** There exists a **node subset** of size $\mathcal{O}(m)$ such that

$$\sup_{\|f\|_{H(K)}\leq 1}\|f-\widetilde{S}^m_{\mathbf{J}}f\|^2_{L_2(D,\varrho)}\leq C\frac{\log(m)}{m}\sum_{k=\lfloor cm\rfloor}^{\infty}\sigma_k^2$$

  $\Longrightarrow$ We produce an additional $\sqrt{\log m}$
- The constants $C, c>0$ are universal and determined explicitly!

Apply Weaver subsampling to RKHS

- Random approach requires $\mathcal{O}(m\log(m))$ samples for a reasonable approximation ($\Longrightarrow$ least squares system matrix $\mathbf{L}_m$ with $m\log m$ rows
- We can "shrink" the matrix $\mathbf{L}_m$ to $\mathcal{O}(m)$ lines
  $\Longrightarrow$ **Weaver subsampling**, see also **Sparsification**, etc.
- Instead of operator $\widetilde{S}_{\mathbf{X}}^m$ we use the operator $\widetilde{S}_{\mathbf{J}}^m$, where $\mathbf{J}$ refers to the relevant samples
- **Consequence:** There exists a **node subset** of size $\mathcal{O}(m)$ such that

$$\sup_{\|f\|_{H(K)}\leq 1} \|f-\widetilde{S}_{\mathbf{J}}^m f\|^2_{L_2(D,\varrho)} \leq C\frac{\log(m)}{m}\sum_{k=\lfloor cm\rfloor}^{\infty}\sigma_k^2$$

  $\Longrightarrow$ We produce an additional $\sqrt{\log m}$
- The constants $C, c>0$ are universal and determined explicitly!

Apply Weaver subsampling to RKHS

- Random approach requires $\mathcal{O}(m\log(m))$ samples for a reasonable approximation ($\Longrightarrow$ least squares system matrix $\mathbf{L}_m$ with $m\log m$ rows
- We can "shrink" the matrix $\mathbf{L}_m$ to $\mathcal{O}(m)$ lines
  $\Longrightarrow$ **Weaver subsampling**, see also **Sparsification**, etc.
- Instead of operator $\widetilde{S}^m_{\mathbf{X}}$ we use the operator $\widetilde{S}^m_{\mathbf{J}}$, where $\mathbf{J}$ refers to the relevant samples
- **Consequence:** There exists a **node subset** of size $\mathcal{O}(m)$ such that

$$\sup_{\|f\|_{H(K)}\leq 1} \|f-\widetilde{S}^m_{\mathbf{J}}f\|^2_{L_2(D,\varrho)} \leq C\frac{\log(m)}{m}\sum_{k=\lfloor cm\rfloor}^{\infty}\sigma_k^2$$

  $\Longrightarrow$ We produce an additional $\sqrt{\log m}$
- The constants $C, c > 0$ are universal and determined explicitly!

Apply Weaver subsampling to RKHS

- Random approach requires $\mathcal{O}(m\log(m))$ samples for a reasonable approximation ($\Longrightarrow$ least squares system matrix $\mathbf{L}_m$ with $m\log m$ rows
- We can "shrink" the matrix $\mathbf{L}_m$ to $\mathcal{O}(m)$ lines
  $\Longrightarrow$ **Weaver subsampling**, see also **Sparsification**, etc.
- Instead of operator $\widetilde{S}^m_{\mathbf{X}}$ we use the operator $\widetilde{S}^m_{\mathbf{J}}$, where $\mathbf{J}$ refers to the relevant samples
- **Consequence:** There exists a **node subset** of size $\mathcal{O}(m)$ such that

$$\sup_{\|f\|_{H(K)}\le 1} \|f-\widetilde{S}^m_{\mathbf{J}}f\|^2_{L_2(D,\varrho)} \le C\frac{\log(m)}{m}\sum_{k=\lfloor cm\rfloor}^{\infty}\sigma_k^2$$

  $\Longrightarrow$ We produce an additional $\sqrt{\log m}$
- The constants $C, c > 0$ are universal and determined explicitly!

Weaver subsampling again

Theorem (Nitzan, Olevskii, Ulanovskii 2016; Limonova, Temlyakov '20; Nagel, Schäfer, T. Ullrich' 20)

Let $k_1, k_2, k_3 > 0$ *and* $\mathbf{u}_1, ..., \mathbf{u}_n \in \mathbb{C}^m$ *with* $\|\mathbf{u}_i\|_2^2 \leq k_1 \frac{m}{n}$ *for all* $i = 1, ..., n$ *and*

$$k_2 \|\mathbf{w}\|_2^2 \leq \sum_{i=1}^{n} |\langle \mathbf{w}, \mathbf{u}_i \rangle|^2 \leq k_3 \|\mathbf{w}\|_2^2 \quad , \quad \mathbf{w} \in \mathbb{C}^m .$$

There always exists $J \subset \{1, \ldots, n\}$ *with* $|J| \leq c_1 \cdot m$ *such that*

$$c_2 \cdot \frac{m}{n} \|\mathbf{w}\|_2^2 \leq \sum_{j \in J} |\langle \mathbf{w}, \mathbf{u}_j \rangle|^2 \leq c_3 \cdot \frac{m}{n} \|\mathbf{w}\|_2^2 \, , \; \mathbf{w} \in \mathbb{C}^m \, ,$$

where $\quad c_1 = 1642 \frac{k_1}{k_2} \, , \quad c_2 = \min\{k_2, (2+\sqrt{2})^2 k_1\} \, , \quad c_3 = 1642 \frac{k_1 k_3}{k_2} \, .$

In its original NOU: $\|\mathbf{u}_i\|_2^2 = n/m$, **tight frame** and no explicit constants

Application: Weaver subsampling

- Denote the row vectors of $\widetilde{\mathbf{L}}_m$ by $\mathbf{u}_1, \ldots, \mathbf{u}_n$:

$$\begin{pmatrix} ---\mathbf{u}_1--- \\ \vdots \\ ---\mathbf{u}_n--- \end{pmatrix} := \widetilde{\mathbf{L}}_m := \begin{pmatrix} \tilde{\eta}_1(\mathbf{x}^1) & \tilde{\eta}_2(\mathbf{x}^1) & \cdots & \tilde{\eta}_m(\mathbf{x}^1) \\ \vdots & \vdots & & \vdots \\ \tilde{\eta}_1(\mathbf{x}^n) & \tilde{\eta}_2(\mathbf{x}^n) & \cdots & \tilde{\eta}_m(\mathbf{x}^n) \end{pmatrix}.$$

- $\left\{\frac{\mathbf{u}_1}{\sqrt{n}}, \ldots, \frac{\mathbf{u}_n}{\sqrt{n}}\right\}$ constitutes a frame in $\mathbb{C}^m$:

$$\frac{1}{2}\|w\|_2^2 \leq \underbrace{\frac{1}{n}\left\|\widetilde{\mathbf{L}}_m w\right\|_2^2}_{\displaystyle \| \atop \displaystyle \sum_{i=1}^{n} \left|\langle w, \frac{\mathbf{u}_i}{\sqrt{n}}\rangle\right|^2} \leq \frac{3}{2}\|w\|_2^2$$

- Further, for all $i \in \{1, \ldots, n\}$: $\left\|\frac{\mathbf{u}_i}{\sqrt{n}}\right\|_2^2 \leq \frac{\tilde{N}(m)}{n} \leq 2 \cdot \frac{m}{n}$

- We can apply Weaver's theorem with $k_1 = 2$, $k_2 = \frac{1}{2}$, $k_3 = \frac{1}{2}$.
- It yields a **subset** $J \subset \{1, \ldots, n\}$ and corresponding sampling points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ with $|J| \in \mathcal{O}(m)$ such that

$$c\|w\|_2^2 \leq \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2 \leq C\|w\|_2^2$$

for the following **submatrix** of $\widetilde{\mathbf{L}}_m$:

$$\widetilde{\mathbf{L}}_{\mathbf{J},m} := \begin{pmatrix} \vdots & \vdots & & \vdots \\ \tilde{\eta}_1(\mathbf{x}^j) & \tilde{\eta}_2(\mathbf{x}^j) & \cdots & \tilde{\eta}_m(\mathbf{x}^j) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{j \in J} = \begin{pmatrix} \vdots \\ - - - \mathbf{u}_j - - - \\ \vdots \end{pmatrix}_{j \in J} .$$

- **Modify sampling recovery operator:** Take (weighted) samples at the points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ (operator $\widetilde{L}_{\mathbf{J}}$) and use (weighted) least squares for reconstruction (operator $(\widetilde{L}_{\mathbf{J},m})^\dagger$), i.e.,

$$\widetilde{S}_{\mathbf{J}}^m := (\widetilde{L}_{\mathbf{J},m})^\dagger \circ \widetilde{L}_{\mathbf{J}}$$

- We can apply Weaver's theorem with $k_1 = 2$, $k_2 = \frac{1}{2}$, $k_3 = \frac{1}{2}$.
- It yields a **subset** $J \subset \{1, \ldots, n\}$ and corresponding sampling points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ with $|J| \in \mathcal{O}(m)$ such that

$$c\|w\|_2^2 \le \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2 \le C\|w\|_2^2$$

for the following **submatrix** of $\widetilde{\mathbf{L}}_m$:

$$\widetilde{\mathbf{L}}_{\mathbf{J},m} := \begin{pmatrix} \vdots & \vdots & & \vdots \\ \tilde{\eta}_1(\mathbf{x}^j) & \tilde{\eta}_2(\mathbf{x}^j) & \cdots & \tilde{\eta}_m(\mathbf{x}^j) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{j \in J} = \begin{pmatrix} \vdots \\ ---\mathbf{u}_j--- \\ \vdots \end{pmatrix}_{j \in J}.$$

- **Modify sampling recovery operator:** Take (weighted) samples at the points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ (operator $\widetilde{L}_{\mathbf{J}}$) and use (weighted) least squares for reconstruction (operator $\left(\widetilde{L}_{\mathbf{J},m}\right)^\dagger$), i.e.,

$$\widetilde{S}_{\mathbf{J}}^m := \left(\widetilde{L}_{\mathbf{J},m}\right)^\dagger \circ \widetilde{L}_{\mathbf{J}}$$

- We can apply Weaver's theorem with $k_1 = 2$, $k_2 = \frac{1}{2}$, $k_3 = \frac{1}{2}$.
- It yields a **subset** $J \subset \{1, \ldots, n\}$ and corresponding sampling points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ with $|J| \in \mathcal{O}(m)$ such that

$$c\|w\|_2^2 \le \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2 \le C\|w\|_2^2$$

for the following **submatrix** of $\widetilde{\mathbf{L}}_m$:

$$\widetilde{\mathbf{L}}_{\mathbf{J},m} := \begin{pmatrix} \vdots & \vdots & & \vdots \\ \tilde{\eta}_1(\mathbf{x}^j) & \tilde{\eta}_2(\mathbf{x}^j) & \cdots & \tilde{\eta}_m(\mathbf{x}^j) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{j \in J} = \begin{pmatrix} \vdots \\ - - - \mathbf{u}_j - - - \\ \vdots \end{pmatrix}_{j \in J} .$$

- **Modify sampling recovery operator:** Take (weighted) samples at the points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ (operator $\widetilde{L}_{\mathbf{J}}$) and use (weighted) least squares for reconstruction (operator $(\widetilde{L}_{\mathbf{J},m})^\dagger$), i.e.,

$$\widetilde{S}_{\mathbf{J}}^m := (\widetilde{L}_{\mathbf{J},m})^\dagger \circ \widetilde{L}_{\mathbf{J}}$$

- We can apply Weaver's theorem with $k_1 = 2$, $k_2 = \frac{1}{2}$, $k_3 = \frac{1}{2}$.
- It yields a **subset** $J \subset \{1, \ldots, n\}$ and corresponding sampling points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ with $|J| \in \mathcal{O}(m)$ such that

$$c\|w\|_2^2 \leq \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2 \leq C\|w\|_2^2$$

for the following **submatrix** of $\widetilde{\mathbf{L}}_m$:

$$\widetilde{\mathbf{L}}_{\mathbf{J},m} := \begin{pmatrix} \vdots & \vdots & & \vdots \\ \tilde{\eta}_1(\mathbf{x}^j) & \tilde{\eta}_2(\mathbf{x}^j) & \cdots & \tilde{\eta}_m(\mathbf{x}^j) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{j \in J} = \begin{pmatrix} \vdots \\ - - - \mathbf{u}_j - - - \\ \vdots \end{pmatrix}_{j \in J} .$$

- **Modify sampling recovery operator:** Take (weighted) samples at the points $\mathbf{J} := (\ldots, \mathbf{x}^j, \ldots)_{j \in J}$ (operator $\widetilde{L}_{\mathbf{J}}$) and use (weighted) least squares for reconstruction (operator $(\widetilde{L}_{\mathbf{J},m})^\dagger$), i.e.,

$$\widetilde{S}_{\mathbf{J}}^m := (\widetilde{L}_{\mathbf{J},m})^\dagger \circ \widetilde{L}_{\mathbf{J}}$$

Error analysis

- For $f \in H$ with $\|f\|_H \leq 1$ we have

$$\|f - \widetilde{S}^m_{\mathbf{J}} f\|^2_{L_2(D,\varrho)} \leq \sigma^2_{m+1} + \frac{\sigma^2_{\max}(\widetilde{\Phi}_{\mathbf{J},m})}{\sigma^2_{\min}(\widetilde{\mathbf{L}}_{\mathbf{J},m})}\,.$$

- Note

$$\text{(i)} \quad \sigma^2_{\max}(\widetilde{\Phi}_{\mathbf{J},m}) \leq \sigma^2_{\max}(\Phi_m)$$

$$\text{(ii)} \quad \sigma^2_{\min}(\widetilde{\mathbf{L}}_{\mathbf{J},m}) \geq c \cdot m \qquad \overset{\text{vs.}}{\longleftrightarrow} \qquad \sigma^2_{\min}(\widetilde{\mathbf{L}}_m) \geq \frac{n}{2}$$

- (ii) is due to

$$c\|w\|_2^2 \leq \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2^2 \leq C\|w\|_2^2$$

- Altogether

$$\|f - \widetilde{S}^m_{\mathbf{J}} f\|^2_{L_2(D,\varrho)} \lesssim \frac{n}{m}\|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho)} \lesssim \log(m)\|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho)}$$

Error analysis

- For $f \in H$ with $\|f\|_H \leq 1$ we have

$$\|f - \widetilde{S}_{\mathbf{J}}^m f\|_{L_2(D,\varrho)}^2 \leq \sigma_{m+1}^2 + \frac{\sigma_{\max}^2(\widetilde{\Phi}_{\mathbf{J},m})}{\sigma_{\min}^2(\widetilde{\mathbf{L}}_{\mathbf{J},m})}\,.$$

- Note

$$\text{(i)} \quad \sigma_{\max}^2(\widetilde{\Phi}_{\mathbf{J},m}) \leq \sigma_{\max}^2(\Phi_m)$$

$$\text{(ii)} \quad \sigma_{\min}^2(\widetilde{\mathbf{L}}_{\mathbf{J},m}) \geq c \cdot m \qquad \overset{\text{vs.}}{\longleftrightarrow} \qquad \sigma_{\min}^2(\widetilde{\mathbf{L}}_m) \geq \frac{n}{2}$$

- (ii) is due to

$$c\|w\|_2^2 \leq \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2^2 \leq C\|w\|_2^2$$

- Altogether

$$\|f - \widetilde{S}_{\mathbf{J}}^m f\|_{L_2(D,\varrho)}^2 \lesssim \frac{n}{m}\|f - \widetilde{S}_{\mathbf{X}}^m f\|_{L_2(D,\varrho)}^2 \lesssim \log(m)\|f - \widetilde{S}_{\mathbf{X}}^m f\|_{L_2(D,\varrho)}^2$$

Error analysis

- For $f \in H$ with $\|f\|_H \leq 1$ we have

$$\|f - \widetilde{S}^m_{\mathbf{J}} f\|^2_{L_2(D,\varrho)} \leq \sigma^2_{m+1} + \frac{\sigma^2_{\max}(\widetilde{\Phi}_{\mathbf{J},m})}{\sigma^2_{\min}(\widetilde{\mathbf{L}}_{\mathbf{J},m})}\,.$$

- Note

$$\text{(i)} \quad \sigma^2_{\max}(\widetilde{\Phi}_{\mathbf{J},m}) \leq \sigma^2_{\max}(\Phi_m)$$

$$\text{(ii)} \quad \sigma^2_{\min}(\widetilde{\mathbf{L}}_{\mathbf{J},m}) \geq c \cdot m \qquad \xleftrightarrow{\text{vs.}} \qquad \sigma^2_{\min}(\widetilde{\mathbf{L}}_m) \geq \frac{n}{2}$$

- (ii) is due to

$$c\|w\|_2^2 \leq \frac{1}{m}\left\|\widetilde{\mathbf{L}}_{\mathbf{J},m} w\right\|_2^2 \leq C\|w\|_2^2$$

- Altogether

$$\|f - \widetilde{S}^m_{\mathbf{J}} f\|^2_{L_2(D,\varrho)} \lesssim \frac{n}{m}\|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho)} \lesssim \log(m)\|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho)}$$

Main result

Theorem (Nagel, Schäfer, T. Ullrich 2020)

Let $H(K)$ be a **separable** *reproducing kernel Hilbert space on a set $D \subset \mathbb{R}^d$ with*

$$\int_D K(\mathbf{x},\mathbf{x})d\varrho_D(\mathbf{x}) < \infty\,.$$

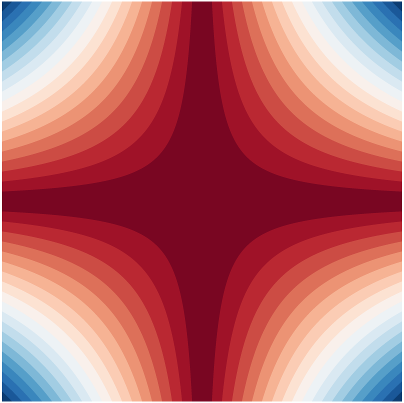
Let further $(\sigma_k)_{k=1}^\infty$ denote the sequence of singular numbers of the Hilbert Schmidt embedding. Then for each $m \geq 2$ there exists a set of sampling nodes $\{\mathbf{x}^1, ..., \mathbf{x}^n\}$ with

$$n \leq 6568 \cdot m$$

such that

$$\sup_{\|f\|_{H(K)}\leq 1} \|f - \widetilde{S}^m_{\mathbf{X}} f\|^2_{L_2(D,\varrho)} \leq 114\frac{\log m}{m}\sum_{k\geq m/2}\sigma_k^2\,. \tag{2}$$

The "smooth" hyperbolic cross



- Special case $p = q = 2$

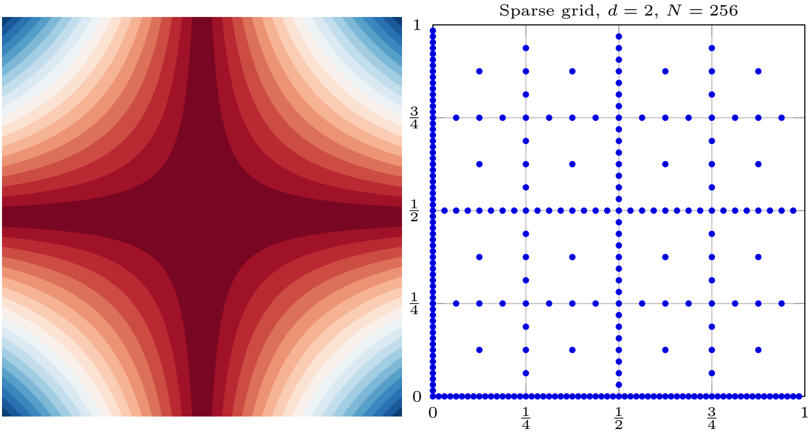
$$\|f\|^2_{H^r_{mix}} := \sum_{k \in \mathbb{Z}} |\hat{f}(k)|^2 \prod_{i=1}^{d} (1+|k_i|^2)^r$$

- Hyperbolic cross projection

$$P_{\mathcal{H}_n} := \sum_{k \in \mathcal{H}_n} \hat{f}(k) e^{2\pi i k x}$$

- **Error:** $\|f - P_{\mathcal{H}_n} f\|_2 \leq n^{-r}$
- **Cost:** $m := \sharp$ grid points in $\mathcal{H}_n$
- **Rate:** $m^{-r} (\log m)^{(d-1)r}$

Sparse grids



- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \qquad (3)$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

An outstanding open problem

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \tag{3}$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \tag{3}$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \tag{3}$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \tag{3}$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

- Function spaces with mixed derivative $H^r_{\text{mix}}(\mathbb{T}^d)$ in $L_2(\mathbb{T}^d)$, $d \in \mathbb{N}, r > 1/2$
- $\sigma_n \asymp n^{-r}(\log n)^{(d-1)r}$, Determinstic approach:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)(r+1/2)} \tag{3}$$

- Probabilistic approach without Weaver subsampling:

$$g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+r}$$

- Improvement over (3) in range $\frac{1}{2} < r < \frac{d-1}{2}$.
- Probbilistic approach with Weaver subsampling:

$$n^{-r}(\log n)^{(d-1)r} \lesssim_{r,d} g_n(H^r_{\text{mix}}, L_2(\mathbb{T}^d)) \lesssim_{r,d} n^{-r}(\log n)^{(d-1)r+1/2}$$

- Improvement over (3) for all $r > \frac{1}{2}$, $d > 2$.

Preasymptotics

Theorem (Kühn, Sickel, U. '15)

Let $r > 0$, $d \in \mathbb{N}$ *and* $1 \le m \le \frac{d}{2}4^d$. *Then*

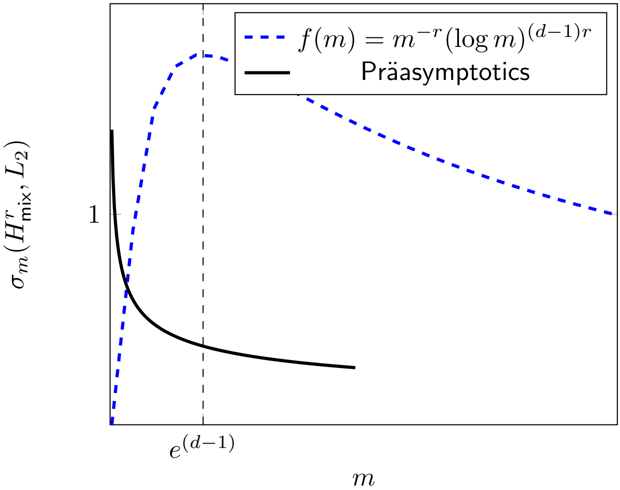
$$\sigma_m(H^r_{\mathrm{mix}}(\mathbb{T}^d), L_2(\mathbb{T}^d)) \le \left(\frac{e^2}{m}\right)^{\frac{r}{2+\log_2 d}}$$

- Extended by **Kühn** to the whole range of m (see also **Krieg '18**)
- Further extension by **Kühn, Sickel, U.** to anisotropic mixed smoothness $\mathbf{r} = (r_1, ..., r_d)$

In case $r > 4 + 2\log_2 d$ the new upper bound on sampling numbers gives a **preasymptotic bound** for sampling

$$\|f - S^m_{\mathbf{X}}\|^2_{L_2} \le \frac{C(4 + 2\log d)}{2r - 4 - 2\log d} \log m \left(\frac{1}{m}\right)^{2r/(2+\log_2 d)}$$

Präasymptotics und asymptotics



Sampling and Approximation Numbers

- n**-th approximation number** a_n measures the minimal worst-case error for recovery from n **linear measurements**

$$a_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{L_1,\ldots,L_n \in H(K)' \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\le 1} \Big\| f - \sum_{i=1}^n L_i(f)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- **Approximation numbers** a_n coincide with the **singular numbers** σ_n of the embedding $\mathrm{Id}: H(K) \to L_2(D,\varrho)$.
- n**-th sampling number** g_n measures the minimal worst-case error for recovery from n **function samples**

$$g_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{\mathbf{x}^1,\ldots,\mathbf{x}^n \in D \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\le 1} \Big\| f - \sum_{i=1}^n f(\mathbf{x}^i)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- Clearly

$$\sigma_n = c_n = a_n\big(H(K), L_2(D,\varrho)\big) \le g_n\big(H(K), L_2(D,\varrho)\big)$$

- n**-th approximation number** a_n measures the minimal worst-case error for recovery from n **linear measurements**

$$a_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{L_1,\ldots,L_n \in H(K)' \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\leq 1} \Big\| f - \sum_{i=1}^{n} L_i(f)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- **Approximation numbers** a_n coincide with the **singular numbers** σ_n of the embedding $\mathrm{Id} : H(K) \to L_2(D,\varrho)$.
- n**-th sampling number** g_n measures the minimal worst-case error for recovery from n **function samples**

$$g_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{\mathbf{x}^1,\ldots,\mathbf{x}^n \in D \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\leq 1} \Big\| f - \sum_{i=1}^{n} f(\mathbf{x}^i)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- Clearly

$$\sigma_n = c_n = a_n\big(H(K), L_2(D,\varrho)\big) \leq g_n\big(H(K), L_2(D,\varrho)\big)$$

- n**-th approximation number** a_n measures the minimal worst-case error for recovery from n **linear measurements**

$$a_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{L_1,\ldots,L_n \in H(K)' \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\le 1} \Big\| f - \sum_{i=1}^n L_i(f)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- **Approximation numbers** a_n coincide with the **singular numbers** σ_n of the embedding $\mathrm{Id} : H(K) \to L_2(D,\varrho)$.
- n**-th sampling number** g_n measures the minimal worst-case error for recovery from n **function samples**

$$g_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{\mathbf{x}^1,\ldots,\mathbf{x}^n \in D \\ \varphi_1,\ldots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)}\le 1} \Big\| f - \sum_{i=1}^n f(\mathbf{x}^i)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- Clearly

$$\sigma_n = c_n = a_n\big(H(K), L_2(D,\varrho)\big) \le g_n\big(H(K), L_2(D,\varrho)\big)$$

- n**-th approximation number** a_n measures the minimal worst-case error for recovery from n **linear measurements**

$$a_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{L_1,\dots,L_n \in H(K)' \\ \varphi_1,\dots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)} \le 1} \Big\| f - \sum_{i=1}^n L_i(f)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- **Approximation numbers** a_n coincide with the **singular numbers** σ_n of the embedding $\mathrm{Id} : H(K) \to L_2(D,\varrho)$.
- n**-th sampling number** g_n measures the minimal worst-case error for recovery from n **function samples**

$$g_n\big(H(K), L_2(D,\varrho)\big) := \inf_{\substack{\mathbf{x}^1,\dots,\mathbf{x}^n \in D \\ \varphi_1,\dots,\varphi_n \in L_2(D,\varrho)}} \sup_{\|f\|_{H(K)} \le 1} \Big\| f - \sum_{i=1}^n f(\mathbf{x}^i)\varphi_i \Big\|_{L_2(D,\varrho)}$$

- Clearly

$$\sigma_n = c_n = a_n\big(H(K), L_2(D,\varrho)\big) \le g_n\big(H(K), L_2(D,\varrho)\big)$$

A new upper bound for sampling numbers

Theorem (Nagel, Schäfer, T. Ullrich 2020)

$H(K)$ as above, K has finite trace, $(\sigma_k)_{k=1}^\infty$ denotes the (non-increasing) sequence of singular numbers of the associated compact embedding $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D,\varrho)$. Then $g_n := g_n(\mathrm{Id}_{K,\varrho})$ satisfies the general bound

$$g_n^2 \le C \frac{\log n}{n} \sum_{k \ge cn} \sigma_k^2$$

with two universal constants $C, c > 0$, which can be specified explicitly.

Improvement over (**Krieg, M. Ullrich '19**):

$$g_n^2 \le C \frac{\log(n)}{n} \sum_{k \ge cn/\log(n)} \sigma_k^2$$

A new upper bound for sampling numbers

Theorem (Nagel, Schäfer, T. Ullrich 2020)

$H(K)$ as above, K has finite trace, $(\sigma_k)_{k=1}^{\infty}$ denotes the (non-increasing) sequence of singular numbers of the associated compact embedding $\mathrm{Id}_{K,\varrho} : H(K) \to L_2(D, \varrho)$. Then $g_n := g_n(\mathrm{Id}_{K,\varrho})$ satisfies the general bound

$$g_n^2 \leq C \frac{\log n}{n} \sum_{k \geq cn} \sigma_k^2$$

with two universal constants $C, c > 0$, which can be specified explicitly.

Improvement over (**Krieg, M. Ullrich '19**):

$$g_n^2 \leq C \frac{\log(n)}{n} \sum_{k \geq cn/\log(n)} \sigma_k^2$$

A new upper bound for sampling numbers

Theorem (Nagel, Schäfer, T. Ullrich 2020)

$H(K)$ as above, K has finite trace, $(\sigma_k)_{k=1}^\infty$ denotes the (non-increasing) sequence of singular numbers of the associated compact embedding $\mathrm{Id}_{K,\varrho}: H(K) \to L_2(D,\varrho)$. Then $g_n := g_n(\mathrm{Id}_{K,\varrho})$ satisfies the general bound

$$g_n^2 \le C \frac{\log n}{n} \sum_{k \ge cn} \sigma_k^2$$

with two universal constants $C, c > 0$, which can be specified explicitly.

Improvement over (**Krieg, M. Ullrich '19**):

$$g_n^2 \le C \frac{\log(n)}{n} \sum_{k \ge cn/\log(n)} \sigma_k^2$$

A further related result...

- Let $D \subset \mathbb{R}^d$ be compact and equipped with probability measure ϱ.
- **Kolmogorov numbers** $(n \in \mathbb{N})$, $\mathbf{F} \subset L_p(D,\varrho)$ centrally symmetric compact subset

$$d_n(\mathbf{F}, L_p) := \inf_{u_1,\ldots,u_n \in L_p} \sup_{f \in \mathbf{F}} \inf_{c_1,\ldots,c_n \in \mathbb{C}} \Big\| f - \sum_{i=1}^n c_i u_i \Big\|_{L_p(D,\varrho)}$$

Theorem (Temlyakov '20)

Let $\mathbf{F}$ be a compact subset of $C(D)$. There exist two constants $C, c > 0$

$$g_{cn}(\mathbf{F}, L_2) \le C d_n(\mathbf{F}, L_\infty) \qquad (n \in \mathbb{N}).$$

- see: **Temlyakov** *On optimal recovery in L_2*, Journ. of Complexity
  Limonova, Temlyakov *On sampling discretization in L_2*, arXiv:2009.10789, 2020
- **Cohen, Migliorati '17**: $\varrho_{bn \log n}(\mathbf{F}, L_2) \lesssim C d_n(\mathbf{F}, L_\infty)$ whp

A further related result...

- Let $D \subset \mathbb{R}^d$ be compact and equipped with probability measure ϱ.
- **Kolmogorov numbers** $(n \in \mathbb{N})$, $\mathbf{F} \subset L_p(D, \varrho)$ centrally symmetric compact subset

$$d_n(\mathbf{F}, L_p) := \inf_{u_1,\ldots,u_n \in L_p} \sup_{f \in \mathbf{F}} \inf_{c_1,\ldots,c_n \in \mathbb{C}} \Big\| f - \sum_{i=1}^{n} c_i u_i \Big\|_{L_p(D,\varrho)}$$

Theorem (Temlyakov '20)

Let $\mathbf{F}$ be a compact subset of $C(D)$. There exist two constants $C, c > 0$

$$g_{cn}(\mathbf{F}, L_2) \le C d_n(\mathbf{F}, L_\infty) \qquad (n \in \mathbb{N}) .$$

- see: **Temlyakov** *On optimal recovery in L_2*, Journ. of Complexity
  Limonova, Temlyakov *On sampling discretization in L_2*, arXiv:2009.10789, 2020
- **Cohen, Migliorati '17**: $\varrho_{bn \log n}(\mathbf{F}, L_2) \lesssim C d_n(\mathbf{F}, L_\infty)$ whp

A further related result...

- Let $D \subset \mathbb{R}^d$ be compact and equipped with probability measure ϱ.
- **Kolmogorov numbers** $(n \in \mathbb{N})$, $\mathbf{F} \subset L_p(D, \varrho)$ centrally symmetric compact subset

$$d_n(\mathbf{F}, L_p) := \inf_{u_1,\ldots,u_n \in L_p} \sup_{f \in \mathbf{F}} \inf_{c_1,\ldots,c_n \in \mathbb{C}} \Big\| f - \sum_{i=1}^{n} c_i u_i \Big\|_{L_p(D,\varrho)}$$

Theorem (Temlyakov '20)

Let $\mathbf{F}$ be a compact subset of $C(D)$. There exist two constants $C, c > 0$

$$g_{cn}(\mathbf{F}, L_2) \le C d_n(\mathbf{F}, L_\infty) \qquad (n \in \mathbb{N}) .$$

- see: **Temlyakov** *On optimal recovery in L_2*, Journ. of Complexity
  Limonova, Temlyakov *On sampling discretization in L_2*, arXiv:2009.10789, 2020
- **Cohen, Migliorati '17**: $\varrho_{bn \log n}(\mathbf{F}, L_2) \lesssim C d_n(\mathbf{F}, L_\infty)$ whp.

A further related result...

- Let $D \subset \mathbb{R}^d$ be compact and equipped with probability measure ϱ.
- **Kolmogorov numbers** $(n \in \mathbb{N})$, $\mathbf{F} \subset L_p(D,\varrho)$ centrally symmetric compact subset

$$d_n(\mathbf{F}, L_p) := \inf_{u_1,\dots,u_n \in L_p} \sup_{f \in \mathbf{F}} \inf_{c_1,\dots,c_n \in \mathbb{C}} \Big\| f - \sum_{i=1}^{n} c_i u_i \Big\|_{L_p(D,\varrho)}$$

Theorem (Temlyakov '20)

Let $\mathbf{F}$ be a compact subset of $C(D)$. There exist two constants $C, c > 0$

$$g_{cn}(\mathbf{F}, L_2) \le C d_n(\mathbf{F}, L_\infty) \qquad (n \in \mathbb{N}).$$

- see: **Temlyakov** *On optimal recovery in L_2*, Journ. of Complexity
  Limonova, Temlyakov *On sampling discretization in L_2*, arXiv:2009.10789, 2020
- **Cohen, Migliorati '17**: $\varrho_{bn \log n}(\mathbf{F}, L_2) \lesssim C d_n(\mathbf{F}, L_\infty)$ whp

A further related result...

- Let $D \subset \mathbb{R}^d$ be compact and equipped with probability measure ϱ.
- **Kolmogorov numbers** $(n \in \mathbb{N})$, $\mathbf{F} \subset L_p(D, \varrho)$ centrally symmetric compact subset

$$d_n(\mathbf{F}, L_p) := \inf_{u_1,\ldots,u_n \in L_p} \sup_{f \in \mathbf{F}} \inf_{c_1,\ldots,c_n \in \mathbb{C}} \Big\| f - \sum_{i=1}^{n} c_i u_i \Big\|_{L_p(D,\varrho)}$$

Theorem (Temlyakov '20)

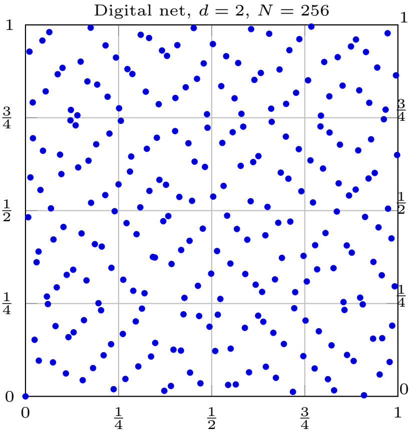
Let $\mathbf{F}$ be a compact subset of $C(D)$. There exist two constants $C, c > 0$

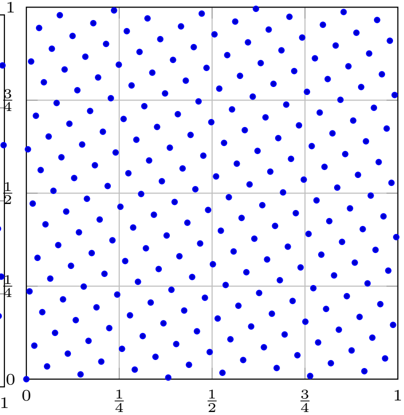
$$g_{cn}(\mathbf{F}, L_2) \le C d_n(\mathbf{F}, L_\infty) \qquad (n \in \mathbb{N}).$$

- see: **Temlyakov** *On optimal recovery in L_2*, Journ. of Complexity
  Limonova, Temlyakov *On sampling discretization in L_2*, arXiv:2009.10789, 2020
- **Cohen, Migliorati '17**: $\varrho_{bn \log n}(\mathbf{F}, L_2) \lesssim C d_n(\mathbf{F}, L_\infty)$ whp

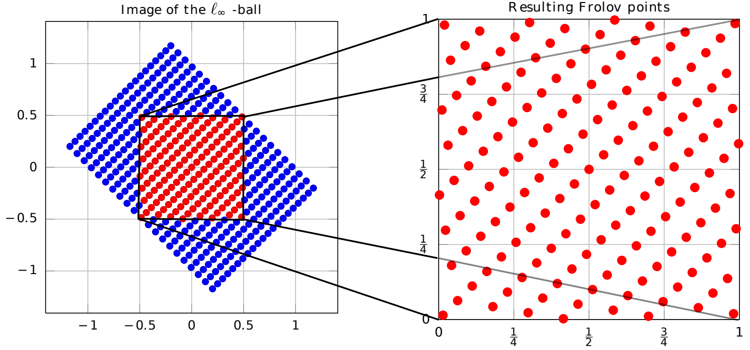
Outlook

QMC nodes


Digital net, $d = 2$, $N = 256$
1
$\frac{3}{4}$
$\frac{1}{2}$
$\frac{1}{4}$
0
0
$\frac{1}{4}$
$\frac{1}{2}$
$\frac{3}{4}$
1


1
$\frac{3}{4}$
$\frac{1}{2}$
$\frac{1}{4}$
0
0
$\frac{1}{4}$
$\frac{1}{2}$
$\frac{3}{4}$
1

Frolov nodes



Kacwin, Oettershagen, M. Ullrich, T.Ullrich

`https://ins.uni-bonn.de/content/software-frolov`

Partial random quadrature nodes

- $H^w(\mathbb{T}^d)$, space with norm $\Big(\sum_{\mathbf{k}\in\mathbb{Z}^d} w(\mathbf{k})^2|\hat{f}(\mathbf{k})|^2\Big)^{1/2} < \infty$

Theorem (Bartel, Kämmerer, Potts, T. Ullrich '21)

Let H^w be as above, $I \subset I' \subset \mathbb{Z}^d$ be frequency index sets and $\mathbf{X} = \{\mathbf{x}^1, ..., \mathbf{x}^M\}$, $(\lambda_1, ..., \lambda_M)$ the nodes / weights of a quadrature rule being exact on $D(I')$. Let further

$$|I| \le \frac{n}{8r\log n} \tag{4}$$

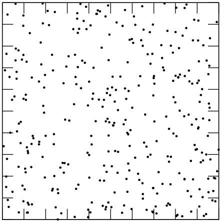
hold true. Drawing $\mathbf{X}_n = (\mathbf{x}^1, \dots, \mathbf{x}^n)$ i.i.d. from $\mathbf{X}$ with respect to the discrete density weigths λ_i, we have with probability larger than $1 - 4n^{1-r}$

$$\sup_{\|f\|_{H^w}\le 1} \Big\| f - S_I^{\mathbf{X}_n} f \Big\|^2_{L_2(\mathbb{T}^d)} \le 5 \sup_{\mathbf{k}\notin I} \frac{1}{w(\mathbf{k})^2} + \frac{7}{|I|}\sum_{\mathbf{k}\notin I'} \frac{1}{w(\mathbf{k})^2} + \sup_{\|f\|_{H^w}\le 1} \Big\| f - S_{I'}^{\mathbf{X}} f \Big\|^2_{L_2(\mathbb{T}^d)}$$

Random nodes

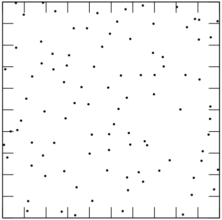
hyperbolic cross with 37 frequencies

$3n \log n$



- 333 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.81631$

$3n$

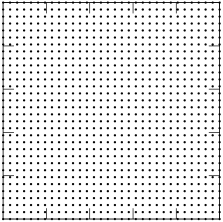


- 95 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.92028$

Equispaced nodes

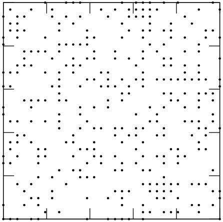
hyperbolic cross with 37 frequencies

n^2



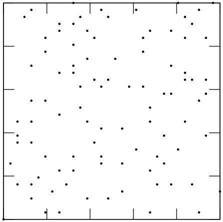
- 1023 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.0000$

$3n \log n$



- 333 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.5884$

$3n$

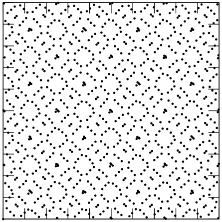


- 95 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.8330$

Sobol nodes

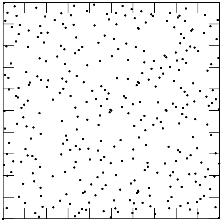
hyperbolic cross with 37 frequencies

n^2



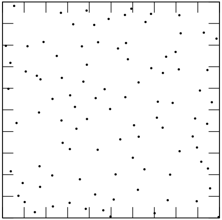
- 1023 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.09794$

$3n\log n$



- 333 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.59935$

$3n$

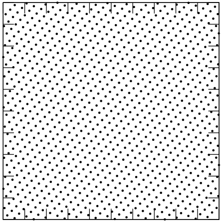


- 95 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.81752$

Frolov nodes

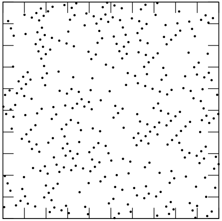
hyperbolic cross with 37 frequencies

n^2



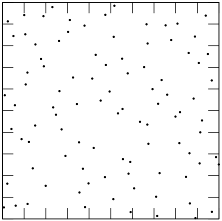
- 1023 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.05439$

$3n \log n$



- 333 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.53356$

$3n$

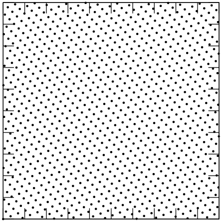


- 95 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.72624$

Fibonacci lattice

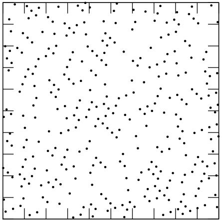
hyperbolic cross with 37 frequencies

n^2



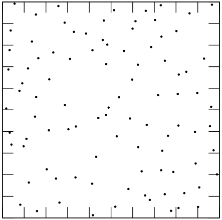
- 987 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.00000$

$3n\log n$



- 325 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.60443$

$3n$



- 94 nodes
- $\|\boldsymbol{F}^*\boldsymbol{F} - \boldsymbol{I}\|_2 \approx 0.83077$

Thank you for your attention!

- D. Krieg and M. Ullrich, *Function values are enough for L_2-approximation*, FoCM, to appear.
- D. Krieg and M. Ullrich, *Function values are enough for L_2-approximation, Part II*, J. Complexity, to appear.
- L. Kämmerer, T. Ullrich, T. Volkmer, *Worst-case recovery guarantees for least squares approximation using random samples*. arXiv:1911.10111, 2019.
- I. Limonova, V. Temlyakov, *On sampling discretization in L_2*. arXiv:2009.10789, 2020.
- M. Moeller, T. Ullrich, *L_2-norm sampling discretization and recovery of functions from RKHS with finite trace*, arXiv:2009.11940, 2020
- N. Nagel, M. Schäfer, T. Ullrich, *A new upper bound for sampling numbers*, FoCM, to appear, arXiv:2010.00327, 2020
- V. Temlyakov, *On optimal recovery in L_2*, arXiv:2010.03103, 2020