Random matrices and L_2-approximation using function values

Mario Ullrich

JKU Linz & MCFAM Moscow

Point Distribution Webinar

Online, May 2021

Motivation

We want to recover/approximate

a function $f : D \to \mathbb{R}$

(or some property of it) up to

a certain error $\varepsilon > 0$,

where f is only known through

some pieces of information.

Motivation

We want to recover/approximate

a function $f\colon D \to \mathbb{R}$?

(or some property of it) up to

a certain error $\varepsilon > 0$, ??

where f is only known through

some pieces of information. ???

During this talk ...

we consider

- a measure space $(D, \mathcal{A}, \mu)$,
- $L_2 = L_2(D, \mathcal{A}, \mu)$: the square-integrable functions w.r.t. μ, and
- a separable metric space $F \hookrightarrow L_2$ of functions on D.

For example:

- $D = [0,1]^d$ or $D = \mathbb{R}^d$ or $D = \mathbb{N}$, with arbitrary μ, and
- F is the unit ball of a separable normed space.

$\big($$F \hookrightarrow L_2$ means here that $\mathrm{id}\colon F \to L_2$, $\mathrm{id}(f) = f$, is injective and compact.$\big)$

Approximation

We want to “compute” an L_2-approximation of $f \in F$ based on a finite (preferably small) number of information, because we ...

- don’t know f and we can only take some measurements, or
- know f, but want to compress it because of computing issues.

What information is allowed,

and how important is this choice?

(The statement “$f \in F$” can be seen as the a priori knowledge about f.)

Information

Information of a function $f \in F$ is given by $L(f)$ for some linear functional $L\colon F \to \mathbb{R}$.

In general, we do not have access to arbitrary $L \in F'$ (=dual of F).

Instead, we have a class of **admissible information** $\Lambda \subset F'$, e.g.,

- certain expectations of f,
- coefficients w.r.t. a given basis,
- **function values:** $f(x)$ for $x \in D$.

Algorithms & error

For information (maps) $L_1, \ldots, L_n \in \Lambda$, we study **linear algorithms**:

$$A_n(f) \;=\; \sum_{i=1}^{n} L_i(f) \cdot \varphi_i$$

for some $\varphi_i \in L_2$. So, A_n is specified by L_i, φ_i.

We want to bound the **worst-case error** over F:

$$e(A_n, F) \;=\; \sup_{f \in F} \left\| f - A_n(f) \right\|_{L_2}.$$

(Several other settings are possible here. Linearity has advantages.)

Minimal worst-case errors

We are interested in the **(linear) sampling numbers**

$$g_n(F) := \inf_{\substack{x_1,\dots,x_n \in D \\ \varphi_1,\dots,\varphi_n \in L_2}} \sup_{f \in F} \left\| f - \sum_{i=1}^{n} f(x_i)\, \varphi_i \right\|_{L_2},$$

i.e., the minimal error that can be achieved with n function values.

As a benchmark, we use the **approximation numbers** (linear width)

$$a_n(F) := \inf_{\substack{L_1,\dots,L_n \in F' \\ \varphi_1,\dots,\varphi_n \in L_2}} \sup_{f \in F} \left\| f - \sum_{i=1}^{n} L_i(f)\, \varphi_i \right\|_{L_2},$$

i.e., the minimal error that can be achieved with arbitrary info.

How good are function values?

The a_n's are well understood, but the g_n's are harder to analyze.

We clearly have

$$a_n(F) \leq g_n(F)$$

if point evaluation $f \mapsto f(x)$ is a continuous linear functional on F.

How large is the difference between g_n and a_n?

Earlier results

Several specific, but only some general bounds were known before.

A negative result [Hinrichs/Novak/Vybíral 2008]

For any $(a_n) \notin \ell_2$, there exist F with $a_n(F) = a_n$ for all n, but

$$g_n(F) \geq \frac{1}{\log\log(n)}.$$

for infinitely many n.

A positive result [Kuo/Wasilkowski/Woźniakowski 2009]

For unit balls of Hilbert spaces H with $a_n(H) \lesssim n^{-\alpha}$, $\alpha > 1/2$, we have

$$g_n(H) \lesssim n^{-\alpha \frac{2\alpha}{2\alpha+1}} \lesssim n^{-\alpha/2}.$$

A very positive result

We now have this general result on the **power of function values**.

Theorem [Krieg/U 2019; U 2020; Krieg/U 2021]

Let $F \hookrightarrow L_2$ be a separable metric space of functions on D, such that point evaluation is continuous on F.

Then, for every $0 < p < 2$, there is a constant $c_p > 0$, depending only on p, such that, for all $n \geq 2$, we have

$$g_N(F) \leq \sqrt{\log n} \left(\frac{1}{n} \sum_{k \geq n} a_k(F)^p \right)^{1/p}$$

for $N \geq c_p \cdot n$.

For unit balls of Hilbert spaces, $p = 2$ also works.[Nagel, Schäfer, T. Ullrich, 2020]

In particular, ...

Corollary

If F is such that

$$a_n(F) \lesssim n^{-\alpha} \log^{\beta}(n)$$

for some $\alpha > 1/2$ and $\beta \in \mathbb{R}$, then we obtain

$$g_n(F) \lesssim n^{-\alpha} \log^{\beta+1/2}(n).$$

Stated differently: If $n \approx (\frac{1}{\varepsilon})^q$, $q < 2$, (arbitrary) infos are enough for an approximation with error $\varepsilon > 0$, then $\left(\frac{\sqrt{\log(1/\varepsilon)}}{\varepsilon}\right)^q$ function values can do the same.

My favorite example

A prominent example:

Sobolev spaces with (dominating) mixed smoothness.

Let $D = \mathbb{T}^d$ be the d-dim. torus, $\mu = \lambda$ the Lebesgue measure on $\mathbb{T}^d$, $1 \le p \le \infty$ and $s \in \mathbb{N}$. We define

$$\mathbf{W}_p^s = \left\{ f \in L_p(\mathbb{T}^d) \colon \|f\|_{\mathbf{W}_p^s} \le 1 \right\},$$

where

$$\|f\|_{\mathbf{W}_p^s} := \left(\sum_{\alpha \in \mathbb{N}_0^d \colon |\alpha|_\infty \le s} \|D^\alpha f\|_p^p \right)^{1/p}.$$

So, $f \in \mathbf{W}_p^s$ implies $D^\alpha f \in L_p$ for all $\alpha \in \mathbb{N}_0^d$ with $\max_i |\alpha_i| \le s$.

My favorite example II

It is known that these well-studied spaces satisfy

- $g_n(\mathbf{W}_p^s) \asymp a_n(\mathbf{W}_p^s)$ for $p < 2$ and all $s > 1/p$.
- $g_n(\mathbf{W}_p^s) \geq a_n(\mathbf{W}_p^s) \asymp n^{-s} \log^{s(d-1)}(n)$ for $p \geq 2$ and $s > 0$.
- $g_n(\mathbf{W}_p^s) \lesssim n^{-s} \log^{(s+1/2)(d-1)}(n)$ for $p \geq 2$ and $s > 1/2$.

All the upper bounds are achieved by sparse grids.[Sickel, T. Ullrich, 2007]

It was the prevalent conjecture that the upper bounds are sharp.

We now have

$$g_n(\mathbf{W}_p^s) \lesssim n^{-s} \log^{s(d-1)+1/2}(n) \quad \text{for } p \geq 2 \text{ and all } s > 1/2.$$

Existence of good points

That is, **sparse grids are not optimal!**

Unfortunately, our general result is only an **existence result**.

In particular, we don't know good sampling points, yet.

However, if we weaken the bound a bit, then ...

- **i.i.d. random points** work with high probability, and
- we know an (to some extent) **explicit algorithm**.

Least squares for function values

It is a classical to study **weighted least squares methods**:

$$A_N(f) = \operatorname*{argmin}_{g \in V_n} \sum_{i=1}^{N} d_i \, |g(x_i) - f(x_i)|^2$$

for some weigths $d_i > 0$, $x_i \in D$ and $V_n = \mathrm{span}\{b_1, \ldots, b_n\} \subset L_2$.

The analysis often boils down to the study of quantities depending on

$$\sum_{k=1}^{n} |b_k(x)|^2 \qquad \text{and} \qquad (f - P_n f)(x).$$

(There are hundreds of results on such methods, mostly for special F.)

Least squares: our approach

To compare $g_n(F)$ and $a_n(F)$, we consider

$$A_N(f) = \operatorname*{argmin}_{g \in V_n} \sum_{i=1}^{N} \frac{|g(x_i) - f(x_i)|^2}{\varrho(x_i)}$$

with $V_n = \mathrm{span}\{b_1, \ldots, b_n\}$, where $\{b_k\}$ **is a "good" basis of** F, and

$$\varrho(x) := \frac{1}{2} \left(\frac{1}{n} \sum_{k \le n} |b_k(x)|^2 + \sum_{k > n} w_k \, |b_k(x)|^2 \right)$$

for some sequence (w_k), s.t. ρ is a μ-density, and choose

$$x_1, \ldots, x_N \overset{\text{iid}}{\sim} \rho \cdot \mathrm{d}\mu.$$

The general result

Theorem [Krieg/U 2021]

Let $F_0 \subset L_2(\mu)$ be a countable set and $x_1, \ldots, x_N \overset{\text{iid}}{\sim} \rho \cdot \mathrm{d}\mu$.

Then, for every $0 < p < 2$, there is a constant $c_p > 0$, depending only on p, such that, for all $n \geq 2$, we have

$$e(A_N, F_0) \leq \left(\frac{1}{n} \sum_{k \geq n} a_k(F_0)^p \right)^{1/p}$$

for $N \geq c_p \, n \log(n)$ with probability at least $1 - \frac{1}{n^2}$.

(For unit balls of Hilbert spaces, $p = 2$ also works.[Krieg/U 2019])

My favorite example III

For the spaces $\mathbf{W}_p^s$ the "good" ONB is given by $\{e^{2\pi i k \cdot} \colon k \in \mathbb{Z}^d\}$, i.e. the Fourier basis. Since $\|b_k\|_\infty \lesssim 1$, we can use $\rho \equiv 1$.

Corollary [Krieg/U 2019, U 2020]

Let $x_1, \dots, x_n$ be independent and uniformly distributed in $\mathbb{T}^d$. Then, for any $s > 1/2$,

$$e(A_n, \mathbf{W}_2^s) \;\lesssim\; a_{\frac{n}{\log n}}(\mathbf{W}_2^s) \;\asymp\; n^{-s} \log^{sd}(n)$$

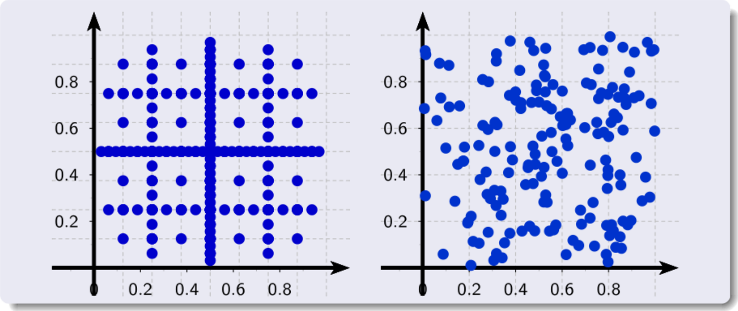
with probability at least $1 - \frac{8}{n^2}$.

Nagel/Schäfer/T. Ullrich 2020: $\quad g_n(\mathbf{W}_2^s) \;\lesssim\; n^{-s} \log^{s(d-1)+1/2}(n)$.

Sparse grids vs. random point sets

$$w.h.p.: \qquad e(A_n, \mathbf{W}_2^s) \lesssim n^{-s} \log^{sd}(n),$$

which is better than sparse grids for $d > 2s + 1$.



What are optimal points?

The proof

The first important insight is that A_N can be written as

$$A_N(f) = \sum_{k=1}^{n} \big(G^+ N(f)\big)_k b_k,$$

where $N\colon F_0 \to \mathbb{R}^n$ with $N(f) = \Big(\varrho(x_i)^{-1/2} f(x_i)\Big)_{i\le N}$ is the **weighted information mapping** and

$G^+ \in \mathbb{R}^{n\times N}$ is the Moore-Penrose inverse of the matrix

$$G = \left(\frac{b_j(x_i)}{\sqrt{\varrho(x_i)}}\right)_{i\le N, j\le n} \in \mathbb{R}^{N\times n}.$$

The proof II

Since A_N is exact on V_n, we obtain

$$\begin{aligned} \|f - A_N f\|_{L_2} &\le \|f - P_n f\|_{L_2} + \|P_n f - A_n f\|_{L_2} \\ &\le a_n + \left\|G^+ N(f - P_n f)\right\|_{\ell_2^n} \\ &\le a_n + \left\|G^+ \colon \ell_2^N \to \ell_2^n\right\| \cdot \|N(f - P_n f)\|_{\ell_2^N} \end{aligned}$$

and hence

$$e(A_N, F_0) \le a_n + s_{\min}(G)^{-1} \sup_{f \in F_0} \|N(f - P_n f)\|_{\ell_2^N},$$

where $s_{\min}$ denotes the smallest singular value.

The proof III

$$e(A_N, F_0) \leq a_n + s_{\min}(G)^{-1} \sup_{f \in F_0} \|N(f - P_n f)\|_{\ell_2^N},$$

We will show that

Fact 1: $s_{\min}(G\colon \ell_2^n \to \ell_2^N)^2 \gtrsim N$

Fact 2: $\sup_{f \in F_0} \|N(f - P_n f)\|_{\ell_2^N}^2 \lesssim n \log n \left(\frac{1}{n} \sum_{k \geq n} a_k^p \right)^{2/p}$

for $N \approx c_p\, n \log(n)$ simultaneously with high probability.

The proof: main tool

Proposition [Oliveira 2010, Mendelson/Pajor 2006]

Let X be a random vector in $\mathbb{C}^k$ with $\|X\|_2 \leq R$ with probability 1, and let $X_1, X_2, \ldots$ be independent copies of X. Additionally, let $E := \mathbb{E}(XX^*)$ satisfy $\|E\| \leq 1$, where $\|E\|$ denotes the spectral norm of E. Then, for all $t \geq \frac{1}{2}$,

$$\mathbb{P}\left(\left\|\sum_{i=1}^{N} X_i X_i^* - N \cdot E\right\| \geq N \cdot t\right) \leq 4N^2 \exp\left(-\frac{N}{32R^2}\, t\right).$$

Note that the bound is dimension-free.

The proof of Fact 1

Let $X_i := \varrho(x_i)^{-1/2}(b_1(x_i), \dots, b_n(x_i))^\top$ with $x_i \sim \rho$. Then, we have

$$\sum_{i=1}^{N} X_i X_i^* \;=\; G^* G \;=\; \left(\sum_{i=1}^{N} \frac{\overline{b_j(x_i)}\, b_k(x_i)}{\varrho(x_i)} \right)_{j,k \le n} \in \mathbb{R}^{n \times n}$$

and $E = \mathbb{E}(XX^*) = \mathrm{diag}(1, \dots, 1)$, i.e., $\|E\| = 1$. Moreover,

$$\|X_i\|_2^2 \;=\; \varrho(x_i)^{-1} \sum_{k \le n} |b_k(x_i)|^2 \;\le\; 2n \;=:\; R^2,$$

since

$$\varrho(x) \;\ge\; \frac{1}{2n} \sum_{k \le n} |b_k(x)|^2.$$

The proof of Fact 1

With $t = \frac{1}{2}$ and $N = \lceil C_1 n \log n \rceil$, we obtain

$$\mathbb{P}\Big(\|G^*G - NE\| \geq \frac{N}{2} \Big) \leq \frac{4}{n^2}$$

if the constant $C_1 > 0$ is large enough. We obtain

$$s_{\min}(G)^2 = s_{\min}(G^*G) \geq s_{\min}(NE) - \|G^*G - NE\| \geq \frac{N}{2}$$

with probability at least $1 - \frac{4}{n^2}$.

The proof of Fact 2: Decomposition

With $I_\ell := \{n2^\ell + 1, \ldots, n2^{\ell+1}\}$, $\ell \geq 0$, and the random matrices

$$\Gamma_\ell := \Big(\varrho(x_i)^{-1/2} b_k(x_i)\Big)_{i\leq N, k\in I_\ell} \in \mathbb{R}^{N\times n2^\ell},$$

and $\hat{f}_\ell := (\langle f, b_k\rangle_{L_2})_{k\in I_\ell}$, we obtain that

$$\|N(f-P_nf)\|_{\ell_2^N} \stackrel{?}{=} \Big\|\sum_{\ell=0}^{\infty} \Gamma_\ell \hat{f}_\ell\Big\|_{\ell_2^N} \leq \sum_{\ell=0}^{\infty} \|\Gamma_\ell \colon \ell_2(I_\ell) \to \ell_2^m\| \, \|\hat{f}_\ell\|_{\ell_2(I_\ell)}$$

$$\leq 2 \sum_{\ell=0}^{\infty} \|\Gamma_\ell \colon \ell_2(I_\ell) \to \ell_2^m\| \;\; a_{n2^{\ell-2}}(F_0)$$

for all $f \in F_0$. The last inequality is ensured by the "good" basis.

The proof of Fact 2: individual blocks

For fixed ℓ, let $X_i := \varrho(x_i)^{-1/2}(b_k(x_i))^\top_{k\in I_\ell}$ with $x_i \sim \rho$. We have

$$\sum_{i=1}^N X_i X_i^* \;=\; \Gamma_\ell^* \Gamma_\ell \;=\; \left(\sum_{i=1}^N \frac{\overline{b_j(x_i)}\, b_k(x_i)}{\varrho(x_i)}\right)_{j,k\in I_\ell} \in \mathbb{R}^{n2^\ell \times n2^\ell}$$

and $E = \mathbb{E}(XX^*) = \mathrm{diag}(1,\dots,1)$, i.e., $\|E\| = 1$. Moreover,

$$\|X_i\|_2^2 \;=\; \varrho(x_i)^{-1} \sum_{k\in I_\ell} |b_k(x_i)|^2 \;\le\; \frac{2}{w_{n2^{\ell+1}}} \;=:\; R^2,$$

since

$$\varrho(x) \;\ge\; \frac{1}{2} \sum_{k\in I_\ell} w_k |b_k(x)|^2 \;\ge\; \frac{w_{n2^{\ell+1}}}{2} \sum_{k\in I_\ell} |b_k(x)|^2.$$

The proof of Fact 2: union bound

With $t \approx \frac{\log(n\ell)}{w_{n2^\ell}\log(n)}$ and $N = \lceil C_1 n \log n \rceil$, we obtain with $\|\Gamma_\ell\|^2 \leq m + \|\Gamma_\ell^* \Gamma_\ell - mE\|$ that

$$\mathbb{P}\left(\|\Gamma_\ell\|^2 \geq C_2\, n \log(n)\, B_\ell^2 \right) \leq \frac{4}{n^2(\ell+1)^2\pi^2}$$

for some $B_\ell \gg \sqrt{\ell\, 2^\ell}$ that is independent of n, N.

We obtain by a union bound that

$$\mathbb{P}\left(\exists \ell \in \mathbb{N}_0 \colon \; \|\Gamma_\ell\|^2 \geq C_2\, n \log(n)\, B_\ell^2 \right) \leq \frac{1}{n^2}.$$

The proof of Fact 2: some calculation

Hence,

$$\|N(f - P_n f)\|_{\ell_2^N} \lesssim n \log(n) \sum_{\ell=0}^{\infty} B_\ell \, a_{n2^\ell}(F_0)$$

for all $f \in F_0$ with probability at least $1 - \frac{1}{n^2}$.

Monotonicity of (a_n) gives

$$\sum_{k \geq n} a_k^p \geq n(2^\ell - 1) \, a_{n2^\ell}^p$$

for $\ell \geq 1$ and thus $a_{n2^\ell} \lesssim 2^{-\ell/p} \left(\frac{1}{n} \sum_{k \geq n} a_k^p \right)^{1/p}$.

We can choose suitable w_k, B_ℓ if $p \in (0, 2)$, which finishes the proof.

The proof of Fact 2: point-wise convergence

It remains to verify $\|N(f - P_n f)\|_{\ell_2^N} \overset{?}{=} \left\| \sum_{\ell=0}^{\infty} \Gamma_\ell \hat{f}_\ell \right\|_{\ell_2^N}$:

We implicitly use

$$(f - P_n f)(x_i) = \sum_{k>n} \hat{f}(k)\, b_k(x_i).$$

Rademacher-Menchov theorem

Let F_0 be **countable** with $\left(\sqrt{\frac{\log(k)}{k}} \cdot a_k(F_0)\right) \in \ell_2$. Then, there is a measurable subset D_0 of D with $\mu(D \setminus D_0) = 0$ such that

$$f(x) = \sum_{k\in\mathbb{N}} \langle f, b_k \rangle_{L_2}\, b_k(x) \qquad \text{for all } x \in D_0 \text{ and } f \in F_0.$$

The proof: From countable to separable

$F \hookrightarrow L_2$ is a separable metric space with cont. point evaluation.

- F contains a countable dense subset F_0
- $\Big\| f - A_N(f) \Big\|_{L_2} \le \Big\| f - g \Big\|_{L_2} + \Big\| g - A_N(g) \Big\|_{L_2} + \Big\| A_N(f-g) \Big\|_{L_2}$
- $U_\delta(f) := \{ g \in F \colon d_F(f,g) < \delta \}$ and $\delta > 0$ small enough
- $g \in F_0 \cap U_\delta(f)\colon \quad \|f - g\|_{L_2} < \varepsilon$ and $|f(x_i) - g(x_i)| < \varepsilon$ (!!!)
- $\Big\| f - A_N(f) \Big\|_{L_2} \le \sup_{g \in F_0} \Big\| g - A_N(g) \Big\|_{L_2} + C\varepsilon$

Hence,

$$e(A_N, F) = e(A_N, F_0) \qquad \text{for every } \underline{\text{linear}} \ A_N.$$

The last step: Downsampling

To finish the proof, we take n "good" out of $n \log n$ random points.

Weaver's theorem [Weaver '04, MSS '15, NOU '16, NSU '20]

There exist constants $c_1, c_2, c_3 > 0$ such that, for all $u_1, \dots, u_N \in \mathbb{C}^n$ such that $\|u_i\|_2^2 \le 2n$ for all $i = 1, \dots, N$ and

$$\frac{1}{2}\|w\|_2^2 \le \frac{1}{N}\sum_{i=1}^{N} |\langle w, u_i\rangle|^2 \le \frac{3}{2}\|w\|_2^2, \qquad w \in \mathbb{C}^n,$$

there is a $J \subset \{1, \dots, m\}$ with $\#J \le c_1 n$ and

$$c_2\,\|w\|_2^2 \le \frac{1}{n}\sum_{i\in J} |\langle w, u_i\rangle|^2 \le c_3\|w\|_2^2, \qquad w \in \mathbb{C}^n.$$

(This is based on the famous solution of the Kadison-Singer problem.)

Finally...

Theorem [Krieg/U 2021]

Let $F \hookrightarrow L_2$ be a separable metric space of functions on D, such that point evaluation is continuous on F.
Then, for every $0 < p < 2$, there is a constant $c_p > 0$, depending only on p, such that, for all $n \geq 2$, we have

$$g_N(F) \leq \sqrt{\log n} \left(\frac{1}{n} \sum_{k \geq n} a_k(F)^p \right)^{1/p}$$

for $N \geq c_p \cdot n$.

Special information

In the above, there's nothing special about function values, and we can do the same for **other classes on information**:

Given a class $\Lambda \subset F'$ of admissible information, let

$$a_n(F,\Lambda) \ := \inf_{\substack{L_1,\dots,L_n\in\Lambda \\ \varphi_1,\dots,\varphi_n\in L_2}} \ \sup_{f\in F} \left\| f - \sum_{i=1}^n L_i(f)\,\varphi_i \right\|_{L_2},$$

be the n-th minimal worst-case error of linear algorithms based on optimal info from Λ.

Special info: The result

Theorem [work in progress]

Let $\Lambda \subset F'$ be such that there exist a measure ν on Λ with

$$\int_\Lambda L(f) \cdot \overline{L(g)} \, \mathrm{d}\nu(L) \; = \; \langle f, g \rangle_{L_2}$$

for all $f, g \in F$.

Then,

$$a_N(F, \Lambda) \; \le \; \sqrt{\log n} \, \left(\frac{1}{n} \sum_{k \ge n} a_k(F)^p \right)^{1/p}$$

for $0 < p < 2$ and $N \ge c_p \cdot n$.

One obtains better bounds for more special info...

Special info: Example

Consider an **arbitrary orthonormal basis**

$$\mathcal{H} = \{h_1, h_2, \dots\} \text{ of } L_2.$$

By choosing ν to be the counting measure, we see

$$\int_\Lambda c(f) \cdot \overline{c(g)}\, \mathrm{d}\nu(c) \;=\; \sum_{i=1}^{\infty} \langle f, h_i \rangle \cdot \overline{\langle g, h_i \rangle} \;=\; \langle f, g \rangle_{L_2}.$$

$\rightsquigarrow$ In this formulation, F does not appear at all.

$\rightsquigarrow$ Your favorite L_2-basis gives almost optimal info if $(a_n) \in \ell_2$.

Special info: The algorithm

For a given class of admissible info $\Lambda \subset F'$, and given $c_1, \dots, c_N \in \Lambda$, let

$$A_N(f) = \operatorname*{argmin}_{g \in V_n} \sum_{i=1}^{N} \frac{|c_i(g) - c_i(f)|^2}{\varrho(c_i)}$$

with

$$\varrho : \Lambda \to \mathbb{R}, \quad \varrho(c) = \frac{1}{2} \left(\frac{1}{n} \sum_{k \le n} |c(b_k)|^2 + \sum_{k > n} w_k |c(b_k)|^2 \right).$$

Final remarks

Open problems:

1. Is the $\sqrt{\log(n)}$-factor needed?
2. Find an explicit construction of such point sets!
3. What are necessary/sufficient conditions?

Note: Lattices don't work. Nets?

$\rightsquigarrow$ We still don't know enough about some of the easiest (general) approximation problems in high dimensions...

Thank you!

How special is optimal information?

One may deduce the following heuristic:

1. For $(a_n) \notin \ell_2$: Optimal information is rare.
2. For $(a_n) \in \ell_2$: (Almost) optimal information is nothing special.

The "good" basis

It is not hard to show that similar holds true for general classes F:

Lemma

There is an orthonormal system $\{b_k \colon k \in \mathbb{N}\}$ *in* L_2 *such that the orthogonal projection* P_n *onto the span* $V_n = \mathrm{span}\{b_1, \dots, b_n\}$ *satisfies*

$$\sup_{f \in F} \|f - P_n f\|_{L_2} \leq 2\, a_{n/4}(F), \qquad n \in \mathbb{N}.$$

- This system is not known in general.
- The '$n/4$' might be problematic for rapidly decaying a_n.

Why mixed smoothness?

Spaces with mixed smoothness are of interest (for numerics) because they ...

- are tensor products of univariate spaces.
- correspond to several concepts of "uniform distribution theory".
- reflect the independence of parameters in high-dimensional models, like medical data, physical measurements etc.
- are proven to be important for the electronic Schrödinger equation. [Yserentant, 2005]

Non-linear algorithms

One might want to consider **arbitrary algorithms**:

$$A_n(f) = \psi\Big(L_1(f), \dots, L_n(f)\Big) \in L_2$$

with some $L_1, \dots, L_n \in F'$ and a (non-linear) mapping $\psi\colon \mathbb{R}^n \to L_2$.

Gelfand width:

$$c_n(F, \Lambda) := \inf_{\substack{\psi\colon \mathbb{R}^n \to L_2 \\ L_1,\dots,L_n \in \Lambda}} \sup_{f \in F} \Big\| f - \psi\Big(L_1(f), \dots, L_n(f)\Big) \Big\|_{L_2}.$$

$$c_n(F) := c_n(F, F')$$

Non-linear algorithms II

Let F be a unit ball of a Banach space.

Several results are known to compare these quantities:

Linear vs. non-linear: $\sup_F \left\{ \frac{a_n(F)}{c_n(F)} \right\} \asymp \sqrt{n}$

Linear vs. non-linear sampling: $\sup_F \left\{ \frac{g_n(F)}{c_n(F,\{\delta_x\})} \right\} \asymp \sqrt{n}$

Lower bound for sampling:

$$g_n(W_1^s([0,1])) \gtrsim c_n(W_1^s([0,1]),\{\delta_x\}) \asymp 1 \text{ for } s < 1.$$

See books of Novak/Wozniakowski 08-12 (Chapter 29), Pinkus etc.

Non-linear algorithms III

Since our result implies

$$g_N(F) \leq \sqrt{\log n} \left(\frac{1}{n} \sum_{k \geq n} \left(\sqrt{k}\, c_k(F) \right)^p \right)^{1/p}$$

for $N \geq c_p \cdot n$, we also know what happens here in the "worst case":

For F a unit ball of a Banach space, we have for $s > 1$

$$n^{-s+1/2} \lesssim \sup\left\{ g_n(F) \colon F \text{ with } c_n(F) \leq n^{-s} \right\} \lesssim \sqrt{\log n} \cdot n^{-s+1/2}$$

and for $s \leq 1$

$$\sup\left\{ g_n(F) \colon F \text{ with } c_n(F) \leq n^{-s} \right\} \asymp 1$$