Widths of convex sets and the power of adaption and randomization

Mario Ullrich
JKU Linz

SIGMA workshop,
October 2024

Framework

We consider the following setting:

- a set of *inputs* $F \subset X$ with a normed space X,
  (often, $F = B_X$ is the unit ball of X)
- a normed space Y, ($\rightarrow$ specifying the error measure)
- a *solution operator* $S\colon X \rightarrow Y$, and
- a class of *admissible information* $\Lambda \subset X' = \{\text{dual space of } X\}$;
  today only $\Lambda = X'$

Goal: *Compute* $S(f)$ for $f \in F$ up to error ε using only info from Λ.

Example: $S\colon X \rightarrow Y$ with $S(f) = f$, i.e., approximation of $f \in X$ in $\|\cdot\|_Y$.

Algorithms

For functionals $c_1, \ldots, c_n \in \Lambda$ (aka information maps), we may use arbitrary *reconstruction mappings* to *approximate* $S \colon X \to Y$ on F:

$$A_n(f) = \varphi\Big(c_1(f), \ldots, c_n(f)\Big) \in Y$$

with some (nonlinear) mapping $\varphi : \mathbb{R}^n \to Y$, and (adaptively chosen) information c_i. We write $A_n = \varphi \circ N_n$, with $N_n = (c_1, \ldots, c_n) \in \Lambda^n$.

We do not care much about φ here and ask the following question:

How much can be gained by choosing the information adaptively and/or randomly?

Algorithms: deterministic

Adaption: information mapping is given recursively by

$$N_n(f) = \Big(N_{n-1}(f),\, L_n(f)\Big),$$

where the choice of the n-th linear functional may depend on the first $n-1$ measurements, i.e., $L_n = L_n(\cdot\,; N_{n-1}(f), L_1, \dots, L_{n-1})$

We denote the set of all such algorithms by $\mathcal{A}_n^{\rm det}(F,Y)$, or just $\mathcal{A}_n^{\rm det}$.

An algorithm is called **non-adaptive** if $N_n = (L_1, \dots, L_n)$, i.e., the same functionals are used for every input.
We denote by $\mathcal{A}_n^{\text{det-non}}$ the corresponding class of algorithms.

Algorithms: randomized

Randomized algorithms are random variables whose realizations are deterministic algorithms.

That is, a randomized algorithm $A_n \colon \Omega \times F \to Y$ is specified by a family of algorithms $(A_n^\omega)_{\omega\in\Omega} \subset \mathcal{A}_n^{\mathrm{det}}(F, Y)$ and a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

- $\mathcal{A}_n^{\mathrm{ran}}(F, Y)$ is the class of all such (possibly adaptive) algorithms
- $\mathcal{A}_n^{\mathrm{ran\text{-}non}}(F, Y)$ is the class of randomized algorithms whose realizations are non-adaptive.

(In general, $\mathcal{A}_n^{\mathrm{det}} \not\subset \mathcal{A}_n^{\mathrm{ran}}$ due to measurability, see the paper.)

Worst-case errors

We define the **worst-case error** for approximating S over F …

- for an algorithm $A_n \in \mathcal{A}_n^{\mathrm{det}} \cup \mathcal{A}_n^{\mathrm{det\text{-}non}}$ by

$$e(A_n, S, F) \ := \ \sup_{f \in F} \left\| S(f) - A_n(f) \right\|_Y .$$

- for an algorithm $A_n \in \mathcal{A}_n^{\mathrm{ran}} \cup \mathcal{A}_n^{\mathrm{ran\text{-}non}}$ by

$$e(A_n, S, F) \ := \ \sup_{f \in F} \mathbb{E} \left\| S(f) - A_n(f) \right\|_Y .$$

(We may omit the Y in $\|\cdot\|_Y$.)

Minimal worst-case errors

The *n*-**th minimal worst-case errors** for approximating S over F:

$$e_n^*(S,F) \ := \inf_{A_n \in \mathcal{A}_n^*} e(A_n, S, F),$$

where $* \in \{\mathrm{det}, \mathrm{det\text{-}non}, \mathrm{ran}, \mathrm{ran\text{-}non}\}$.

Known: these minimal errors can be quite different...

Can we say something about the maximal difference?

In the following, we assume $S \in \mathcal{L}(X,Y)$, i.e., S **is linear & bounded**.

Known results in special cases

Theorem [Bakhvalov '71, Gal/Micchelli '80, Traub/Wozniakowski '80]

For every convex & symmetric $F \subset X$ and $n \in \mathbb{N}$, we have

$$e_n^{\text{det-non}}(S,F) \;\leq\; 2\, e_n^{\text{det}}(S,F).$$

Theorem [Novak '92]

Let H, G be Hilbert spaces and $S \in \mathcal{L}(H,G)$. For all $n \in \mathbb{N}$, we have

$$e_{2n}^{\text{det-non}}(S,B_H) \;\leq\; 2\, e_n^{\text{ran}}(S,B_H).$$

Theorem [Novak '95]

For every convex $F \subset X$ and $n \in \mathbb{N}$, we have

$$e_n^{\text{det-non}}(S,F) \;\leq\; 4(n+1)^2\; e_n^{\text{det}}(S,F).$$

Is adaption useless for symmetric sets?

If F is convex and symmetric, adaption does not help for deterministic algorithms. It was open for a long time whether the same holds for randomized algorithms.

This problem was recently solved by Stefan Heinrich who considered *parametric integration* using function values as Λ.
For $\Lambda = X'$:

Theorem [Kunsch/Novak/Wnuk '24, Kunsch/Wnuk '24]

Let $S\colon \ell_1^m \to \ell_\infty^m$, $S(f) = f$, and $B_1 := B_{\ell_1^m}$ for suitable $m = m(n)$, then

$$e_n^{\mathrm{ran}}(S, B_1) \lesssim \frac{\log n}{n}\, e_n^{\mathrm{ran\text{-}non}}(S, B_1).$$

This (clearly) also implies $e_n^{\mathrm{ran}}(S, B_1) \lesssim_{\log} \frac{1}{n} \cdot e_n^{\mathrm{det}}(S, B_1)$.

Main result

Theorem [Krieg/Novak/U '24]

Let X, Y be Banach spaces and $S \in \mathcal{L}(X, Y)$.

For every convex $F \subset X$ and $n \in \mathbb{N}$, we have

$$e_{2n}^{\text{det-non}}(S, F) \;\le\; 12\, n^{3/2} \left(\prod_{k<n} e_k^{\text{ran}}(S, F) \right)^{1/n}.$$

In special cases, the following improvements hold:

1. if F is symmetric, we can replace $n^{3/2}$ with n,
2. if Y is a Hilbert space, we can replace $n^{3/2}$ with n,
3. if F is symmetric and Y a Hilbert space, replace $n^{3/2}$ with $n^{1/2}$,
4. if X is a Hilbert space and F its unit ball, we can replace $n^{3/2}$ with $n^{1/2}$ if we additionally replace the index $2n$ with $4n$.

Main result II

It might be easier to digest in the following form:

Theorem [Krieg/Novak/U '24]

For every convex $F \subset X$, $n \in \mathbb{N}$ and $\alpha > 0$, we have

$$e_{2n}^{\text{det-non}}(S,F) \;\leq\; C_\alpha\, n^{-\alpha+3/2} \cdot \sup_{k<n} \Big((k+1)^\alpha\, e_k^{\text{ran}}(S,F) \Big)\,,$$

where $C_\alpha \leq 12^{\alpha+1}$.

If F is convex and symmetric (like $F = B_X$), then

$$e_{2n}^{\text{det-non}}(S,F) \;\leq\; C_\alpha\, n^{-\alpha+1} \cdot \sup_{k<n} \Big((k+1)^\alpha\, e_k^{\text{ran}}(S,F) \Big)\,.$$

Again, more improvements for X or Y being Hilbert spaces.

State of the art

These bounds on the **maximal gain in the rate of convergence**, together with some specific examples, imply the following table:

Gain from	$\mathcal{A}_n^{\text{det-non}}$		$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{det}}$
to / for	$\mathcal{A}_n^{\text{det}}$	$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{ran}}$	
F convex+symmetric	0	$\left[\frac{1}{2}, 1\right]$	1	1
F convex	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$

(We ignore logarithmic factors.)

n-widths and s-numbers

A crucial tool are inequalities between n-**widths of sets** and/or **s-numbers of operators** (in the sense of Pietsch), for which we provide a common generalization.

First, we define the **Gelfand numbers** of $S \in \mathcal{L}(X,Y)$ on $F \subset X$ by

$$c_n(S,F) := \inf_{L_1,\dots,L_n \in X'} \sup_{\substack{f,g\in F:\\ L_k(f)=L_k(g)}} \frac{1}{2} \left\| S(f) - S(g) \right\|.$$

For $S = \mathrm{id}_X$, i.e., identity on X, these are the *Gelfand widths* of F, and for $F = B_X$ these are the *Gelfand numbers* of S.

Two other "widths"...

The **Bernstein numbers** of $S \in \mathcal{L}(X,Y)$ on $F \subset X$:

$$b_n(S,F) := \sup_{\substack{V\subset X \text{ affine} \\ \dim(V)=n+1}} \sup_{g\in F\cap V} \inf_{f\in V\cap(X\setminus F)} \|S(f)-S(g)\|,$$

i.e., the largest $(n+1)$-dim. ball in F w.r.t. norm $\|f\|_S := \|S(f)\|_Y$.

The **Hilbert numbers** of $S \in \mathcal{L}(X,Y)$ on $F \subset X$:

$$h_n(S,F) := \sup\Big\{c_n(CSA, B_{\ell_2}) \colon\ C \in \mathcal{L}(Y,\ell_2) \text{ with } \|C\| \le 1, \\ A \in \mathcal{L}(\ell_2, X) \text{ and } x \in F \text{ with } A(B_{\ell_2}) + x \subset F\Big\},$$

i.e., Gelfand numbers of the "most difficult Hilbertian sub-problem".

Bounds between s-numbers

For $F = B_X$, i.e., for s-numbers, a lot is known, mainly due to Pietsch (who recently passed away). For example, from [Pietsch '74]:

$$h_n(S, B_X) \;\leq\; b_n(S, B_X) \;\leq\; c_n(S, B_X).$$

The following reverse inequality was essentially proved by Pietsch in the 1980s; better constant and accessible proof were recently observed.

Theorem [Pietsch '80, U '24]

For all $S \in \mathcal{L}(X, Y)$ and $n \in \mathbb{N}$,

$$c_n(S, B_X) \;\leq\; n \cdot \Big(\prod_{k<n} h_k(S, B_X) \Big)^{1/n}.$$

We extended this to general F, which requires an additional $\sqrt{n}$.

Widths versus minimal errors

It remains to connect minimal errors to the different widths.

First, Gelfand numbers are basically the minimal errors for $\mathcal{A}_n^{\text{det-non}}$:

Theorem [Traub/Wozniakowski '80]

For every $S \in \mathcal{L}(X, Y)$, $F \subset X$ and $n \in \mathbb{N}_0$, we have

$$c_n(S, F) \leq e_n^{\text{det-non}}(S, F) \leq 2\, c_n(S, F).$$

Second, Bernstein numbers are lower bounds on errors for $\mathcal{A}_n^{\text{ran}}$:

Theorem [Kunsch '17]

For every $S \in \mathcal{L}(X, Y)$, convex $F \subset X$ and $n \in \mathbb{N}_0$, we have

$$e_n^{\text{ran}}(S, F) \geq \frac{1}{30}\, b_{2n}(S, F).$$

Other widths

The **Kolmogorov numbers** of S on F:

$$d_n(S,F) := \inf_{\substack{M\subset Y\\ \dim(M)\le n}} \sup_{f\in F} \inf_{g\in M} \|S(f)-g\|_Y \,,$$

i.e., the error of best-approximation on an optimal subspace.

($\to$ no direct relation to algorithms)

Theorem [Pietsch '80]

Let $S \in \mathcal{L}(X,Y)$ and F convex and symm. with $b_n(S,F) \asymp b_{2n}(S,F)$. Then, for all $n \in \mathbb{N}$,

$$d_n(S,F) \lesssim n \cdot b_n(S,F).$$

This solves an old problem of Mityagin and Henkin (1963), at least for regularly decaying b_n. (This seems to have gone unnoticed...)

Other widths II

There are also *non-linear widths* that can be bounded by this method.

The **manifold widths** of $F \subset X$ are defined by

$$\delta_n(F) := \inf_{\substack{N \in C(X,\mathbb{R}^n) \\ \varphi \in C(\mathbb{R}^n, X)}} \sup_{f \in F} \left\| f - \varphi(N(f)) \right\|,$$

where $C(X, Y)$ denotes the class of continuous maps from X to Y.

It is known that $\delta_n(F) \gtrsim b_n(F)$. [DeVore et al '89]

Theorem [KNU '24]

For all convex F, we have

$$c_{2n}(\mathrm{id}_X, F) \lesssim n^{-\alpha+3/2} \cdot \sup_{k<n} (k+1)^{\alpha} \, \delta_k(F),$$

Also holds with "d_{2n}", and smaller exponents under assumptions.

Teaser: Adaption with continuous info

What if we allow **adaptive, continuous information**, i.e.,
$N_n(f) = \big(N_{n-1}(f),\, L_n(f)\big)$ with $L_n \in C(X,\mathbb{R})$ chosen adaptively?

Denote the corresponding minimal errors by $e_n^{\text{cont-ada}}(S,F)$.

Very recently, we obtained the following (surprising?) result:

Theorem [KNU '25?]

Let X, Y be Banach spaces, Y separable and $S \in \mathcal{L}(X,Y)$.
Then, for all $F \subset X$ and $n \in \mathbb{N}$, we have

$$e_{n+1}^{\text{cont-ada}}(S,F) \lesssim d_{2^n}(S,F).$$

E.g., every $x \in \mathbb{R}^m$ can be recovered up to an arbitrary small ε using only $\sim \log(m)$ adaptive measurements. Discussion?!?

What about ...

- bounds for individual n? (i.e., without geometric mean)
- the maximal gain for convex sets? (Is it > 1?)
- restricted classes of linear information?
- non-linear S?

A particularly interesting question on randomization:

Open problem

Is there some $S \in \mathcal{L}(X, Y)$ and $\alpha > 1/2$ with

$$e_n^{\text{ran-non}}(S, B_X) \;\leq\; n^{-\alpha} \cdot e_n^{\text{det-non}}(S, B_X)?$$

Thank you!

Gain from	$\mathcal{A}_n^{\text{det-non}}$		$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{det}}$
to / for	$\mathcal{A}_n^{\text{det}}$	$\mathcal{A}_n^{\text{ran-non}}$	$\mathcal{A}_n^{\text{ran}}$	
F convex+symmetric	0	$\left[\frac{1}{2}, 1\right]$	1	1
F convex	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[\frac{1}{2}, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$	$\left[1, \frac{3}{2}\right]$