How many (noisy) measurements are needed to learn a vector?

Mario Ullrich

Inst. Analysis & Dep. Quantum Computing
JKU Linz

Albert's 60^{th} birthday
Paris, June 2025

The problem

Assume that you want to **recover an arbitrary vector** $x \in \mathbb{R}^m$, up to some error $\varepsilon > 0$ in some norm $\|\cdot\|$, where $m \in \mathbb{N}$ can be large.

You know nothing about x, you can only compute certain measurements $\lambda_1(x), \ldots, \lambda_n(x) \in \mathbb{R}$ that you can choose.

How many measurements do you need?

(We do not assume a bound on a norm of x yet.)

Linear information

Less than m linear measurements are useless.

Proof:

Let $N = (\lambda_1, \ldots, \lambda_n)$ with $n < m$, where the λ_k are linear functionals.

Then, for $y = N(x)$, there is a whole affine subspace V of $\mathbb{R}^m$ with $\dim(V) \geq m - n$ such that $N(v) = y$ for all $v \in V$.

So no matter how you choose your approximation $\hat{x} = \Phi(y) \in \mathbb{R}^m$, you may be arbitrarily far away from the true value of x.

Continuous information

The same is true for any continuous measurement map $N\colon \mathbb{R}^m \to \mathbb{R}^n$.
This follows from the Borsuk-Ulam theorem:

Borsuk-Ulam theorem (∼1930)

For any $N \in C(\mathbb{R}^m, \mathbb{R}^n)$ with $n < m$ and $R > 0$, there are x and $\tilde{x}$ at distance $2R$ such that $N(x) = N(\tilde{x})$.

That is, you cannot distinguish between these vectors and hence cannot guarantee an error less than R.

So, again, **less than m continuous measurements are useless**.

Adaptive information

But why should we fix the measurement maps λ_k in advance?

It is probably better to use the already obtained information $\lambda_1(x),\ldots,\ \lambda_{k-1}(x)$ and choose λ_k based on it.

We call such measurements **adaptive**.

(Precisely, we allow $\lambda_k(x) = \lambda_k(x\,;\lambda_1(x),\ldots,\lambda_{k-1}(x),\lambda_1,\ldots,\lambda_{k-1})$.)

Adaptive linear information

"Unfortunately", the fact remains that

less than m linear measurements are useless,

also if we choose them adaptively.

Proof:

As before, for $y = N(x)$, there is a whole affine subspace $V \subset \mathbb{R}^m$ such that $\lambda_j(v) = y_j$ for all $j \leq n < m$ and $v \in V$.

Thus, for any $v \in V$, you would have chosen the same functionals λ_j and obtained the same information $y = N(v)$.

Hence, you cannot distinguish between all the elements of V.

Adaptive continuous information

The remaining case:

Can we achieve something with less than m adaptively chosen continuous measurements?

History: When we started to work on this problem with David Krieg and Erich Novak during the summer 2024, we studied the "simplest example" $x \in \{x \in \mathbb{R}^3 \colon \|x\|_2 \le 1\}$, i.e., $m = 3$, and $n = 2$. Starting with the functional $\ell_1(x) = x_1 - |x_3|$, one can improve over all nonadaptive methods (which have worst case error 1).

Surprisingly, $\log(m)$ measurements are enough

Theorem [Krieg/Novak/U '25]

Let $m \in \mathbb{N}$ and $\varepsilon > 0$. The algorithm R_m^ε described below uses at most $n(m) := \lceil \log_2(m) \rceil + 1$ adaptive, 1-Lipschitz-continuous measurements and satisfies for all $x \in \mathbb{R}^m$ that

$$\|x - R_m^\varepsilon(x)\| \leq \varepsilon.$$

This implies

$$\|R_m^\varepsilon(x) - R_m^\varepsilon(y)\| \leq \|x - y\| + 2\varepsilon,$$

which might be considered as some kind of *stability*.

Again: There cannot be a continuous algorithm! (Borsuk-Ulam)

The algorithm

The algorithm is based on a **coloring** of $\mathbb{R}^m$ in the following way:

Consider a partition

$$\mathbb{R}^m = \bigcup_{i \in \mathbb{N}} D_i,$$

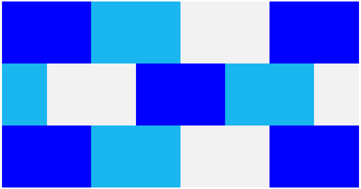
a (coloring) map $t\colon \mathbb{N} \to \{1, 2, \ldots, m+1\}$, and some $c > 0$ with:

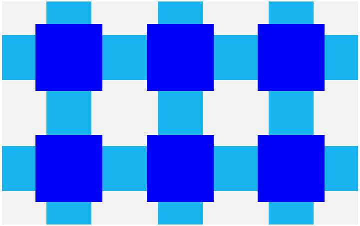
1. $\mathrm{diam}(D_i) \leq 1$ for each i;
2. if $i \neq j$ and $t(i) = t(j)$ then $\mathrm{dist}(D_i, D_j) \geq c$,

where diameter (diam) and distance (dist) are with respect to the given norm on $\mathbb{R}^m$.

Colorings of $\mathbb{R}^m$ with separated colors

The following illustrations show two colorings of the plane with three colors. The second is easily generalized to higher dimensions.





Less than $m+1$ colors do not work.

This is related to the **Nagata dimension** of $\mathbb{R}^m$.

The algorithm: First stage and bisection

To find an ε-approximation of $x \in \mathbb{R}^m$ it is enough to find an index i^* such that x is in εD_{i^*}. Define by

$$I_r := \{i \in \mathbb{N}\colon t(i) = r\} \quad \text{and} \quad E_r = \bigcup_{j \in I_r} \varepsilon D_j$$

the **set of points with color** r in the $\underline{\varepsilon\text{-scaled}}$ partition.

A continuous measurement of the form $\lambda_J(x) = \operatorname{dist}\Big(x, \bigcup_{j \in J} E_j\Big)$ with $J \subset \{1, \ldots, m+1\}$, tells us whether x **has any of the colors in** J**.**

We use $n = \lceil \log_2(m+1) \rceil = n(m) - 1$ such functionals and **bisection to find a color of** x, i.e., $r^* = t(i^*)$ with $x \in \overline{E}_{r^*}$.

Algorithm: Second stage

Now we can determine a correct index i^* with $x \in \varepsilon\overline{D}_{i^*}$ using any continuous functional λ^* for which the images $\lambda^*(\varepsilon\overline{D_i})$ for $i \in I_{r^*}$ are pairwise disjoint. An example is given by

$$\lambda^*(x) = \max_{i \in I_{r^*}} \left\{ \frac{c\varepsilon}{2i} - \operatorname{dist}(x, \varepsilon D_i) \right\}.$$

By construction, x is in the closure of εD_{i^*} with $i^* = \frac{c\varepsilon}{2\lambda^*(x)}$.

The output $R_m^\varepsilon(x)$ of the algorithm can be any element from εD_{i^*}.

Discussion

Did we really construct a clever algorithm for the recovery of $x \in \mathbb{R}^m$ if m is large?

We made two basic assumptions:

1. If $\lambda \colon \mathbb{R}^m \to \mathbb{R}$ is 1-Lipschitz, then we can compute values $\lambda(x)$, i.e., all Lipschitz functionals are admissible as information.
2. The information can be chosen adaptively, i.e., λ_{k+1} may depend on the (already computed) values $y_i = \lambda_i(x)$ for $i = 1, 2, \ldots, k$.

Adaption is widespread and can be easily implemented on a computer. The use of arbitrary Lipschitz measurements is more problematic.

Power of adaption for continuous information

We want to **approximate** $f \in F$ **in the norm of** Y**,** where $F \subset Y$ is subset of (general) metric space Y. We allow **algorithms** $A_n \colon F \to Y$ of the form

$$A_n(f) = \Phi(\lambda_1(f), \ldots, \lambda_n(f))$$

with adaptively chosen continuous measurements $\lambda_i \colon F \to \mathbb{R}$ and an arbitrary reconstruction map $\Phi \colon \mathbb{R}^n \to Y$.

We consider the **minimal worst-case error** of algorithms that use at most n continuous measurements, i.e.,

$$e_n^{\text{cont}}(F) := \inf_{A_n} \sup_{f \in F} d_Y\Big(f, A_n(f)\Big),$$

where the infimum ranges over all adaptive A_n as above.

Power of adaption for continuous information II

We compare with the **manifold widths** of F (in Y), i.e.,

$$\delta_n(F) := \inf_{\substack{N \in C(F,\mathbb{R}^n) \\ \Phi \in C(\mathbb{R}^n, Y)}} \sup_{f \in F} d_Y\big(f, \Phi(N(f))\big).$$

$\rightsquigarrow$ minimal errors of **non-adaptive "continuous algorithms"**

Theorem [KNU '25]

Let $F \subset Y$ be a subset of a metric space Y and $n \geq 2$. Then,

$$e_n^{\text{cont}}(F) \leq \delta_{2n-2}(F).$$

Noisy information

What if measurements are only known up to some noise $\delta > 0$?

Precisely, $(y_1, \ldots, y_n) \in \mathbb{R}^n$ is the information about $f \in F$ which satisfies

$$\Big| y_i - \lambda_i(f; y_1, \ldots, y_{i-1}) \Big| \le \delta, \qquad 1 \le i \le n,$$

for some functionals $\lambda_i(\,\cdot\,; y_1, \ldots, y_{i-1}) \colon F \to [-1, 1]$ and $\delta < 1$.

(Interpretation: *ideal* measurements λ_i; machine precision δ; normalization $[-1, 1]$)

We denote by $e_n^{?}(F, \delta)$ with $? \in \{\text{lin}, \text{cont}, \text{arb}\}$ the minimal error of **arbitrary algorithms using ? measurements that are noisy**.

Linear noisy information

There is lot of work and books on **noisy linear measurements**.

Here, we only state

$$e_n^{\mathrm{lin}}(F,\delta) \;\geq\; \delta \qquad \text{for all } n \in \mathbb{N}.$$

That is, for fixed $\delta > 0$,

there cannot be a (possibly non-linear) algorithm based on noisy linear measurements with error going to zero.

Continuous noisy information

For noisy continuous measurements $\lambda_i\colon\ F \to [-1,1]$, we obtain an upper bound with the **entropy numbers** of $F \subset Y$:

$\varepsilon_n(F) := \inf\{\varepsilon > 0\colon\ F$ can be covered by 2^n balls (in Y) of radius $\varepsilon\}$.

Theorem [Krieg/Novak/Plaskota/U '25]

Let $F \subset Y$ be a subset of a metric space Y, $n \in \mathbb{N}$ and $\delta < 1$. Then,

$$e_n^{\mathrm{cont}}(F,\delta) \le \varepsilon_n(F).$$

- upper bound is independent of $\delta < 1$ (?!)
- $e_n^{\mathrm{cont}}(F,\delta) = \varepsilon_0(F)$ for $\delta \ge 1$
- We did not find a "good" lower bound for $e_n^{\mathrm{cont}}(F,\delta)$.

Arbitrary noisy information

We consider **arbitrary (possibly non-continuous) measurements**.

Here, the minimal errors are even **characterized by the ε_n:**

Theorem [Krieg/Novak/Plaskota/U '25]

Let $F \subset Y$ be a subset of a metric space Y, $n \in \mathbb{N}$ and $\delta < 1$.

Then,

$$\varepsilon_{n(k_\delta+1)}(F) \;\leq\; e_n^{\mathrm{arb}}(F,\delta) \;\leq\; \varepsilon_{nk_\delta}(F)$$

with

$$k_\delta := \lceil \log_2(1/\delta + 1) \rceil - 1.$$

$\rightsquigarrow$ gain of nonlinear adaptive measurements is limited for fixed $\delta > 0$

An example

We illustrate these results by the example of approximating vectors from the unit ball $F = B_p^m$ of ℓ_p^m in the norm of $Y = \ell_q^m$.

We know

$$\varepsilon_n(B_p^m, \ell_q^m) \asymp \left(\frac{\log(m/n+1)}{n}\right)^{\frac{1}{p}-\frac{1}{q}}$$

for $\log(m) \leq n \leq m$ and $p \leq q$.

We will only discuss $p = 2$ and $q = \infty$.

An example II

We denote the number of measurements for ℓ_∞-error $\varepsilon \in (0,1)$ by

$$n^?(\varepsilon,\delta) := \inf\left\{n\colon\ e_n^?(B_2^m,\delta) \le \varepsilon\right\}.$$

An example II

We denote the number of measurements for ℓ_∞-error $\varepsilon \in (0,1)$ by

$$n^{?}(\varepsilon,\delta) \;:=\; \inf\left\{n\colon\; e_n^{?}(B_2^m,\delta) \le \varepsilon\right\}.$$

We have that

- $n^{\mathrm{lin}}(\varepsilon,\delta) \;=\; \infty \quad$ for $\varepsilon < \delta$,

An example II

We denote the number of measurements for ℓ_∞-error $\varepsilon \in (0,1)$ by

$$n^{?}(\varepsilon,\delta) \;:=\; \inf\left\{n\colon\; e_n^{?}(B_2^m,\delta) \le \varepsilon\right\}.$$

We have that

- $n^{\mathrm{lin}}(\varepsilon,\delta) \;=\; \infty \quad$ for $\;\varepsilon < \delta$,
- $n^{\mathrm{lin}}(\varepsilon,\delta) \;\gtrsim\; n^{\mathrm{lin}}(\varepsilon,0) \;\gtrsim\; m \quad$ for $\;\varepsilon > \delta$,

An example II

We denote the number of measurements for ℓ_∞-error $\varepsilon \in (0,1)$ by

$$n^?(\varepsilon,\delta) \;:=\; \inf\left\{n\colon\; e_n^?(B_2^m,\delta) \le \varepsilon\right\}.$$

We have that

- $n^{\mathrm{lin}}(\varepsilon,\delta) \;=\; \infty \qquad \text{for} \;\; \varepsilon < \delta,$
- $n^{\mathrm{lin}}(\varepsilon,\delta) \;\gtrsim\; n^{\mathrm{lin}}(\varepsilon,0) \;\gtrsim\; m \qquad \text{for} \;\; \varepsilon > \delta,$
- $n^{\mathrm{cont}}(\varepsilon,0) \;\le\; \lceil \log_2(m+1) \rceil \qquad \text{for all } \varepsilon > 0,$

An example II

We denote the number of measurements for ℓ_∞-error $\varepsilon \in (0,1)$ by

$$n^?(\varepsilon,\delta) \;:=\; \inf\left\{n\colon\; e_n^?(B_2^m,\delta) \le \varepsilon\right\}.$$

We have that

- $n^{\mathrm{lin}}(\varepsilon,\delta) \;=\; \infty \quad$ for $\;\varepsilon < \delta$,
- $n^{\mathrm{lin}}(\varepsilon,\delta) \;\gtrsim\; n^{\mathrm{lin}}(\varepsilon,0) \;\gtrsim\; m \quad$ for $\;\varepsilon > \delta$,
- $n^{\mathrm{cont}}(\varepsilon,0) \;\le\; \lceil \log_2(m+1) \rceil \qquad$ for all $\varepsilon > 0$,
- $\log_{1/\delta}(m)\cdot\varepsilon^{-2} \;\lesssim\; n^{\mathrm{cont}}(\varepsilon,\delta) \;\lesssim\; \log_2(m)\cdot\varepsilon^{-2}$

Final remarks

- For $Y = \ell_q^m$, $q \in \{1, \infty\}$, the functionals are **piecewise linear**.
- Neural networks...
- For ℓ_2-approximation in $F = B_1^m$, we have

$$n^{\mathrm{lin}}(\varepsilon, \delta) \approx n^{\mathrm{cont}}(\varepsilon, \delta) \qquad \text{for} \quad \varepsilon < \delta.$$

A particularly interesting question:

Open problem

What about **other classes of measurements?**

(just norms; convex, homogeneous, smooth)

Thank you, Albert!

Thank you, Albert!

Lipschitz noisy information

Let us consider **Lipschitz-continuous measurements**.

(The result depends on the Lipschitz constant.)

Theorem [Krieg/Novak/Plaskota/U '25]

For any $\delta \cdot \varepsilon_0(F) < 1$, we have

$$\frac{\delta}{L} \;\leq\; e_n^{\mathrm{Lip}}(F,\delta) \;\leq\; \varepsilon_n(F) + \frac{n\,\delta}{L \cdot \varepsilon_0(F)},$$

where e_n^{Lip} is the n-th minimal error with L-Lipschitz functionals.

In particular,

$$e_n^{\mathrm{cont}}(F,\delta) \leq \varepsilon_n(F).$$