Randomized approximation of finite sequences – adaption vs. non-adaption

Robert J. Kunsch
(joint work with Erich Novak and Marcin Wnuk)

MCM 2023 – Special Session:
On the power of iid information for (non-linear) approximation

Paris, 28 juin 2023

The problem

Consider sequence space embeddings:

$$\ell_p^m \hookrightarrow \ell_q^m, \qquad 1 \leq p, q \leq \infty$$

Approximate $\boldsymbol{x} \in \mathbb{R}^m = \ell_p^m$ using information

$$N\colon \mathbb{R}^m \to \mathbb{R}^n, \qquad m \gg n \text{ (at least } m \geq 2n)$$

- via **deterministic** algorithms $A_n = \phi \circ N$
  $\rightsquigarrow$ minimal worst-case error

$$e^{\text{det}}(n, \ell_p^m \hookrightarrow \ell_q^m) := \inf_{A_n} \sup_{\|\boldsymbol{x}\|_p \leq 1} \|A_n(\boldsymbol{x}) - \boldsymbol{x}\|_q$$

- via **randomized** algorithms $A_n = (\phi^\omega \circ N^\omega)_{\omega \in \Omega}$

$$e^{\text{ran}}(n, \ell_p^m \hookrightarrow \ell_q^m) := \inf_{A_n} \sup_{\|\boldsymbol{x}\|_p \leq 1} \mathbb{E}\,\|A_n(\boldsymbol{x}) - \boldsymbol{x}\|_q$$

Information maps and adaption

$$N\colon \ell_p^m \to \mathbb{R}^n$$

non-adaptive information, representable as matrix $N \in \mathbb{R}^{n\times m}$,

$$\boldsymbol{y} = N\boldsymbol{x} = (L_1(\boldsymbol{x}), \ldots, L_n(\boldsymbol{x})), \qquad L_1, \ldots, L_n \in \ell_p'$$

adaptive information $\boldsymbol{y} = N(\boldsymbol{x})$ where

$$y_i := L_i(f; y_1, \ldots, y_{i-1}), \qquad L_i(\,\cdot\,; y_1, \ldots, y_{i-1}) \in \ell_p'$$

iid information = special non-adaptive randomized information, most prominently Gaussian information with

$$L_i(\boldsymbol{x}) := \sum_{j=1}^{m} g_{ij} x_j, \qquad g_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

Information maps and adaption

$$N\colon \ell_p^m \to \mathbb{R}^n$$

non-adaptive information, representable as matrix $N \in \mathbb{R}^{n\times m}$,

$$\boldsymbol{y} = N\boldsymbol{x} = (L_1(\boldsymbol{x}), \ldots, L_n(\boldsymbol{x})), \qquad L_1, \ldots, L_n \in \ell_p'$$

adaptive information $\boldsymbol{y} = N(\boldsymbol{x})$ where

$$y_i := L_i(f; y_1, \ldots, y_{i-1}), \qquad L_i(\,\cdot\,; y_1, \ldots, y_{i-1}) \in \ell_p'$$

iid information = special non-adaptive randomized information, most prominently Gaussian information with

$$L_i(\boldsymbol{x}) := \sum_{j=1}^m g_{ij} x_j, \qquad g_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

ℓ_2-approximation of summable sequences

$$\ell_1^m \hookrightarrow \ell_2^m$$

- deterministic rate [Kashin 1977, Garnaev & Gluskin 1984],

$$e^{\det}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \sqrt{\frac{\log \frac{m}{n}}{n}}, \quad m \geq 2n$$

  non-linear reconstruction based on iid info $\boldsymbol{y} = N\boldsymbol{x}$ (Gaussian)

$$A_n(\boldsymbol{x}) = \underset{\boldsymbol{z}:\, N\boldsymbol{z}=\boldsymbol{y}}{\operatorname{argmin}} \|\boldsymbol{z}\|_1$$

  (achieves desired worst-case error rate with high probability)

- improvements in randomized setting?

$$\frac{1}{\sqrt{n}} \overset{\text{[Heinrich 1992]}}{\preceq} e^{\text{ran}}(n, \ell_1^m \hookrightarrow \ell_2^m), \qquad m \geq 2n$$

ℓ_2-approximation of summable sequences

$$\ell_1^m \hookrightarrow \ell_2^m$$

- deterministic rate [Kashin 1977, Garnaev & Gluskin 1984],

$$e^{\det}\left(n, \ell_1^m \hookrightarrow \ell_2^m\right) \preceq \sqrt{\frac{\log \frac{m}{n}}{n}}, \quad m \geq 2n$$

  non-linear reconstruction based on iid info $\boldsymbol{y} = N\boldsymbol{x}$ (Gaussian)

$$A_n(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{z}\colon N\boldsymbol{z}=\boldsymbol{y}} \|\boldsymbol{z}\|_1$$

  (achieves desired worst-case error rate with high probability)

- improvements in randomized setting?

$$\frac{1}{\sqrt{n}} \overset{\text{[Heinrich 1992]}}{\preceq} e^{\text{ran}}\left(n, \ell_1^m \hookrightarrow \ell_2^m\right), \qquad m \geq 2n$$

Uniform approximation of square-summable sequences

$$\ell_2^m \hookrightarrow \ell_\infty^m$$

- deterministic approximation hard:

$$e^{\mathrm{det}}(n, \ell_2^m \hookrightarrow \ell_\infty^m) \overset{\text{[Smolyak 1965]}}{\asymp} 1\,, \qquad m \geq 2n$$

- randomized for $m \geq 2n$:

$$\sqrt{\frac{\log n}{n}} \overset{\text{[Heinrich 1992]}}{\preceq} e^{\mathrm{ran}}(n, \ell_2^m \hookrightarrow \ell_\infty^m) \overset{\text{[Mathé 1991]}}{\preceq} \sqrt{\frac{\log m}{n}}$$

linear reconstruction based on iid info $\boldsymbol{y} = N\boldsymbol{x}$ (Gaussian)

$$A_n(\boldsymbol{x}) = \frac{1}{n} N^\top \boldsymbol{y}$$

Uniform approximation of square-summable sequences

$$\ell_2^m \hookrightarrow \ell_\infty^m$$

- deterministic approximation hard:

$$e^{\mathrm{det}}(n, \ell_2^m \hookrightarrow \ell_\infty^m) \overset{\text{[Smolyak 1965]}}{\asymp} 1\,, \qquad m \geq 2n$$

- randomized for $m \geq 2n$:

$$\sqrt{\frac{\log n}{n}} \overset{\text{[Heinrich 1992]}}{\preceq} e^{\mathrm{ran}}(n, \ell_2^m \hookrightarrow \ell_\infty^m) \overset{\text{[Mathé 1991]}}{\preceq} \sqrt{\frac{\log m}{n}}$$

linear reconstruction based on iid info $\boldsymbol{y} = N\boldsymbol{x}$ (Gaussian)

$$A_n(\boldsymbol{x}) = \frac{1}{n} N^\top \boldsymbol{y}$$

Uniform approximation of summable sequences

$$\ell_1^m \hookrightarrow \ell_\infty^m$$

- optimal deterministic rate achieved with linear algorithms,

$$e^{\det}(n, \ell_1^m \hookrightarrow \ell_\infty^m) \preceq \sqrt{\frac{\log \frac{m}{n}}{n}}, \quad m \geq 2n$$

- combining non-linear reconstruction for $\ell_1^m \hookrightarrow \ell_2^m$ with randomized linear method for $\ell_2^m \hookrightarrow \ell_\infty^m$ [Heinrich 1992]

$$\frac{\sqrt{\log n}}{n} \preceq e^{\mathrm{ran}}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \frac{\sqrt{(\log \frac{m}{n})(\log m)}}{n}, \quad m \geq 2n$$

Gaussian iid info feeds both stages, hence, the whole method

Uniform approximation of summable sequences

$$\ell_1^m \hookrightarrow \ell_\infty^m$$

- optimal deterministic rate achieved with linear algorithms,

$$e^{\mathrm{det}}(n, \ell_1^m \hookrightarrow \ell_\infty^m) \preceq \sqrt{\frac{\log \frac{m}{n}}{n}}, \quad m \geq 2n$$

- combining non-linear reconstruction for $\ell_1^m \hookrightarrow \ell_2^m$ with randomized linear method for $\ell_2^m \hookrightarrow \ell_\infty^m$ [Heinrich 1992]

$$\frac{\sqrt{\log n}}{n} \preceq e^{\mathrm{ran}}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \frac{\sqrt{(\log \frac{m}{n})(\log m)}}{n}, \quad m \geq 2n$$

Gaussian iid info feeds both stages, hence, the whole method

Summary of known results

Known results:

- best known rates achieved with iid info (i.e. non-adaptive)
- lower bounds [Heinrich 1992] without m-dependence

New results:

- new lower bounds reflecting dependence on m (for non-adaptive setting, including iid info)
- reduced m-dependence via adaptive info (not iid!)

Summary of known results

Known results:

- best known rates achieved with iid info (i.e. non-adaptive)
- lower bounds [Heinrich 1992] without m-dependence

New results:

- new lower bounds reflecting dependence on m (for non-adaptive setting, including iid info)
- reduced m-dependence via adaptive info (not iid!)

A new lower bound

Theorem (K, Novak, Wnuk 2023*)

For sufficiently large $m \in \mathbb{N}$ there exist constants $c_0, \varepsilon_0 > 0$ such that for

$$n \leq c_0 \sqrt{\log m}$$

we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, \ell_1^m \hookrightarrow \ell_\infty^m) \geq \varepsilon_0 \,.$$

Proof idea: Switch to average case setting, i.e. random input:

$$\boldsymbol{X} = \frac{2}{3} \boldsymbol{e}_I + \frac{1}{12n} P \boldsymbol{Z} \,,$$

with random index $I \sim \mathcal{U}\{1, \ldots, m\}$, Gaussian $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_m)$, and projection P onto $2n$ randomly picked coordinates.
Then, with high probability, $\|\boldsymbol{X} - \frac{2}{3}\boldsymbol{e}_I\|_1 \leq \frac{1}{3}$, especially $\|\boldsymbol{X}\|_1 \leq 1$.

A new lower bound

Theorem (K, Novak, Wnuk 2023*)

For sufficiently large $m \in \mathbb{N}$ there exist constants $c_0, \varepsilon_0 > 0$ such that for

$$n \leq c_0 \sqrt{\log m}$$

we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, \ell_p^m \hookrightarrow \ell_q^m) \geq \varepsilon_0 \,.$$

Proof idea: Switch to average case setting, i.e. random input:

$$\boldsymbol{X} = \frac{2}{3} \boldsymbol{e}_I + \frac{1}{12n} P \boldsymbol{Z} \,,$$

with random index $I \sim \mathcal{U}\{1, \ldots, m\}$, Gaussian $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_m)$, and projection P onto $2n$ randomly picked coordinates.
Then, with high probability, $\|\boldsymbol{X} - \frac{2}{3}\boldsymbol{e}_I\|_1 \leq \frac{1}{3}$, especially $\|\boldsymbol{X}\|_1 \leq 1$.

A new lower bound

Theorem (K, Novak, Wnuk 2023*)

For sufficiently large $m \in \mathbb{N}$ there exist constants $c_0, \varepsilon_0 > 0$ such that for

$$n \leq c_0 \sqrt{\log m}$$

we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, \ell_1^m \hookrightarrow \ell_\infty^m) \geq \varepsilon_0 \,.$$

Proof idea: Switch to average case setting, i.e. random input:

$$\boldsymbol{X} = \frac{2}{3}\boldsymbol{e}_I + \frac{1}{12n} P\boldsymbol{Z}\,,$$

with random index $I \sim \mathcal{U}\{1, \ldots, m\}$, Gaussian $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_m)$, and projection P onto $2n$ randomly picked coordinates.

Then, with high probability, $\|\boldsymbol{X} - \frac{2}{3}\boldsymbol{e}_I\|_1 \leq \frac{1}{3}$, especially $\|\boldsymbol{X}\|_1 \leq 1$.

A new lower bound

Theorem (K, Novak, Wnuk 2023*)

For sufficiently large $m \in \mathbb{N}$ there exist constants $c_0, \varepsilon_0 > 0$ such that for

$$n \leq c_0 \sqrt{\log m}$$

we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, \ell_1^m \hookrightarrow \ell_\infty^m) \geq \varepsilon_0 \,.$$

Proof idea: Switch to average case setting, i.e. random input:

$$\boldsymbol{X} = \frac{2}{3}\boldsymbol{e}_I + \frac{1}{12n} P \boldsymbol{Z} \,,$$

with random index $I \sim \mathcal{U}\{1, \ldots, m\}$, Gaussian $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, I_m)$, and projection P onto $2n$ randomly picked coordinates. Then, with high probability, $\|\boldsymbol{X} - \frac{2}{3}\boldsymbol{e}_I\|_1 \leq \frac{1}{3}$, especially $\|\boldsymbol{X}\|_1 \leq 1$.

Approximability of operators

With $m \to \infty$, we conclude **non-approximability** of ∞-dim. sequence space embeddings and non-compact operators in general:

Corollary (K, Novak, Wnuk 2023*)

There exists $\varepsilon_0 > 0$ such that for all $n \in \mathbb{N}$ we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, \ell_1 \hookrightarrow \ell_\infty) \geq \varepsilon_0 \,.$$

Theorem (K, Novak, Wnuk 2023*)

Let $S\colon F \to G$ be a non-compact linear operator between Banach spaces F, G. Then there exists $\varepsilon_S > 0$ such that for all $n \in \mathbb{N}$ we have the non-adaptive Monte Carlo lower bound

$$e^{\mathrm{ran,nonada}}(n, S) \geq \varepsilon_S \,.$$

Adaptive Monte Carlo methods

Theorem (K, Novak, Wnuk 2023*)

For large $m \in \mathbb{N}$, adaptive Monte Carlo methods yield the upper bound

$$e^{\mathrm{ran,ada}}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \sqrt{\frac{\log\log\frac{m}{n}}{n}}\,.$$

Idea: For an error smaller than $\varepsilon/2 > 0$, it suffices to identify the $k = \lceil 16\varepsilon^{-2} \rceil$ largest entries of $\boldsymbol{x}$ (heavy hitters).
Woodruff et al. (2011, 2019) adaptively identify the heavy hitters with high probability using $n \asymp k \log\log\frac{m}{k}$ samples by sequentially narrowing down the potential locations of the heavy hitters. □

This upper bound also applies for $\ell_1^m \hookrightarrow \ell_\infty^m$.

Adaptive Monte Carlo methods

Theorem (K, Novak, Wnuk 2023*)

For large $m \in \mathbb{N}$, adaptive Monte Carlo methods yield the upper bound

$$e^{\mathrm{ran,ada}}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \sqrt{\frac{\log\log\frac{m}{n}}{n}}\,.$$

Idea: For an error smaller than $\varepsilon/2 > 0$, it suffices to identify the $k = \lceil 16\varepsilon^{-2} \rceil$ largest entries of $\boldsymbol{x}$ (heavy hitters).
Woodruff et al. (2011, 2019) adaptively identify the heavy hitters with high probability using $n \asymp k \log\log\frac{m}{k}$ samples by sequentially narrowing down the potential locations of the heavy hitters. $\square$

This upper bound also applies for $\ell_1^m \hookrightarrow \ell_\infty^m$.

Adaptive Monte Carlo methods

Theorem (K, Novak, Wnuk 2023*)

For large $m \in \mathbb{N}$, adaptive Monte Carlo methods yield the upper bound

$$e^{\mathrm{ran,ada}}(n, \ell_1^m \hookrightarrow \ell_2^m) \preceq \sqrt{\frac{\log\log\frac{m}{n}}{n}}\,.$$

Idea: For an error smaller than $\varepsilon/2 > 0$, it suffices to identify the $k = \lceil 16\varepsilon^{-2} \rceil$ largest entries of $\boldsymbol{x}$ (heavy hitters).
Woodruff et al. (2011, 2019) adaptively identify the heavy hitters with high probability using $n \asymp k \log\log\frac{m}{k}$ samples by sequentially narrowing down the potential locations of the heavy hitters. $\square$

This upper bound also applies for $\ell_1^m \hookrightarrow \ell_\infty^m$.

Gap between non-adaptive and adaptive Monte Carlo

Non-adaptive randomization (including methods with iid info) can be worse than adaptive randomization **for large problems**:

Theorem (K, Novak, Wnuk 2023*)

For $m = \lceil C \exp(an^2) \rceil$ *with suitable* $C, a > 0$*, we obtain the gap*

$$\frac{e^{\mathrm{ran,ada}}(n, \ell_1^m \hookrightarrow \ell_2^m)}{e^{\mathrm{ran,nonada}}(n, \ell_1^m \hookrightarrow \ell_2^m)} \preceq \sqrt{\frac{\log n}{n}}\,.$$

Heinrich (2022/23) showed a similar gap of order $\frac{\log n}{\sqrt{n}}$ for parametric integration and standard info (function evaluation).

Here: Larger gap but for general info (linear functionals)

Gap between non-adaptive and adaptive Monte Carlo

Non-adaptive randomization (including methods with iid info) can be worse than adaptive randomization **for large problems**:

Theorem (K, Novak, Wnuk 2023*)

For $m = \lceil C \exp(an^2) \rceil$ with suitable $C, a > 0$, we obtain the gap

$$\frac{e^{\mathrm{ran,ada}}(n, \ell_1^m \hookrightarrow \ell_2^m)}{e^{\mathrm{ran,nonada}}(n, \ell_1^m \hookrightarrow \ell_2^m)} \preceq \sqrt{\frac{\log n}{n}}.$$

Heinrich (2022/23) showed a similar gap of order $\frac{\log n}{\sqrt{n}}$ for parametric integration and standard info (function evaluation).

Here: Larger gap but for general info (linear functionals)

Merci pour votre attention.

Des questions?

References I

S. A. Smolyak.
On optimal restoration of functions and functionals of them.
Dissertation, Moscow State University, 1965, in Russian.

B. S. Kashin.
Diameters of some finit-dimensional sets and classes of smooth functions.
Mathematics of the USSR Izvestija, 11(2):317–333, 1977, English translation.

A. YU. Garnaev, E. D. Gluskin.
On widths of a Euclidean ball.
Soviet Mathematics Doklady, 30:200–204, 1984, English translation.

References II

P. Mathé.
Random approximation of Sobolev embeddings.
Journal of Complexity, 7(3):261–281, 1991.

S. Heinrich.
Lower bounds for the complexity of Monte Carlo function approximation.
Journal of Complexity, 8(3):277–300, 1992.

Kh. Do Ba, P. Indyk, E. Price, D. P. Woodruff.
Lower bounds for sparse recovery.
in Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, 2010.

References III

P. Indyk, E. Price, D. P. Woodruff.
On the power of adaptivity in sparse recovery.
IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS), 285–294, 2011.

E. Price, D. P. Woodruff.
Lower bounds for adaptive sparse recovery.
SODA '13: Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms, 652–663, 2013.

Yi Li, V. Nakos, D. P. Woodruff
On low-risk heavy hitters and sparse recovery schemes
Preprint, 2019.

References IV

W. Johnson.
A universal non-compact operator
Colloquium Mathematicum 23:267–268, 1971.

S. Heinrich.
Randomized complexity of parametric integration and the role of adaption I. Finite Dimensional Case.
Preprint, 2023.

S. Heinrich.
Randomized complexity of vector-valued approximation
Preprint, 2023.