

Math to power industry project

Edgar, Igor, Symon, Thomas
Department of Mathematics
Simon Fraser University
Canada

September 25, 2021

Known anisotropy:

- Use least squares (standard L2 loss)
- Can optimize the samples via Migliorati, Cohen idea
- DNN architectures will be smaller, therefore beneficial

Unknown anisotropy:

- Use compressed sensing via weighted/unweighted ℓ^1
- Larger DNN architectures

Exponential rates: as in MSML

- Note that with these we can use the preconditioning trick to get $\exp(-\gamma \tilde{m}^{1/d})$ (as opposed to $1/(2d)$ when sampling from the uniform measure). This also works with unweighted ℓ^1

Algebraic rates: new item to add!

Fast algorithms: new stuff

1 Items to investigate

1) Better DNN architectures

- Can we reduce the size of the Opschoor et al ReLU construction?
- They also consider polynomial units – check that
- Investigate the Daws and Webster approach – does that yield smaller DNNs? <https://arxiv.org/abs/1905.10457> <https://arxiv.org/abs/1912.02302>
- Also look at <https://arxiv.org/pdf/1903.05858.pdf>

Sebastian/Ben to dig into this first.

2) Chebyshev polynomials – we need an analogous version of the Opschoor result for these (the result in their paper only considers Legendre)

Sebastian/Ben to dig into this first.

3) Infinite-dimensional case

- I think this is do-able, either via the approach of Rauhut–Schwab or my alternative approach
- To do: also investigate anchored sets (see Cohen, Migliorati, Nobile paper – is that another approach)

Simone and Ben to work on inf dim CS stuff.

After that, we then dig into the Opschoor and Zech, Opschoor, Schwab and Zech papers, along with the Hermann, Schwab and Zech paper.

4) Noncompact domains, e.g. \mathbb{R}^d with Hermite

- The tricky thing here is the DNN approximation of the Hermite polynomial. Is this known? Usually, DNNs approximations only work on compact intervals (although the Hermite polynomials decay exponentially outside a certain window, so perhaps it is ok).

We agree this is probably too much for this paper. But we'll still think about solving this for future work.

2 Items for the future, i.e. to mention in the conclusion

Banach-valued problems

Sebastian/Nick: figure out a convincing parametric PDE problem that is naturally formulated in a Banach space. If you can sell Ben on it, then we can think about it.

CNNs, ResNets, non ReLU activation functions. All work in practice (check this). Can we have any theory for them. Note: this is mentioned in the Kutyniok et al parametric PDE theory paper.

Add to conclusion we provide an upper bound on DNN size. This may well be too large in practice. Can one prove a lower bound?

Discussion – Dec 15, 2020

We discuss splitting into two papers:

Paper 1: algorithms for Hilbert-valued function approximation, via CS. This has all the technical CS theory, plus the PDI algorithm, etc. Includes code for the methods. Contributions:

- Novel, exponentially-convergent algorithm
- Full approximation statement for HV function approximation via CS with an algorithm
- Infinite-dimensional case as well (hopefully)

Nick: focus on the PDI implementation, both the vanilla $1/n$ algorithm and the exponentially-restarted.

Paper 2: DNN approximation. Basically, has the main theorem on DNNs, but also has lots of cool experiments comparing with the method of Paper 1.

Sebastian/Nick/Simone: think about which parametric PDEs to show in this paper. Let's write a list and then decide which ones we're going to implement for the paper.

Are the ones from Geist good to use? What about parametric domain problems? Or time dependent problems? It might also be nice here to do some dimensionality studies, as in Geist et al. Can we make some more concrete takeaways from this? Also, some problems where CS with polynomials does not do well. We could also even do the basic harmonic oscillator problem.