2nd report on “Interpolation of High-Dimensional Data”

by Thomas C. H. Lux et al.

I thank the authors for having taken into account my previous suggestions. The manuscript has improved, and I think it is a valuable and informative reading for many practitioners. However quite some work still has to be done to improve its readability and precision: therefore I suggest a further round of major reviews before the manuscript gets accepted. I list below here the comments that I found while reading.

1. page 2: “Splines in particular are well understood as an interpolation technique”. So do piecewise linear and Lagrangian interpolation schemes. Please fix the sentence and add references. Any advanced textbook of numerical analysis will do.
2. page 2: A high-dimensional interpolation survey is not complete without mentioning the results in sparse grids approximation. There are 30 years of literature on this, with rigorous convergence estimates in arbitrary dimensions, that show how this method can defeat (to a certain extent) the dimensional scaling challenge. Take a look e.g. at [1] and more recent advances such as [2, 4]. The main “problem” with them is that they do not work right away for scattared data.
3. page 2: “generic uniform bounds are largely unobtainable for regression techniques on arbitrary approximation problems”: not entirely true. Bounds for high dimensional regression techniques using multi variate orthogonal polynomials can be found e.g. in [3].
4. page 3: “that is not to say nonlinear models are absent, but nonlinearities are often either preconceived or model pairwise interactions between dimensions at most”. Again not true for sparse grids and projection on orthogonal polynomials (cf. references already mentioned). These methods can reasonably handle interactions between 5-10 variables simultaneously.
5. page 4: Definition of MARS. How many iterations, i.e., what is the range of j?
6. page 4/5: Definition of MLP. Mention that the choice of the network architecture is crucial; use “iterations” instead of “epochs”.
7. page 5, Delaunay. Add a picture of a mesh that satisfies the Delaunay criterion and one that doesn’t. Given a collection of scattered points, can one always produce a Delaunay triangulation out of them?
8. page 6, Modified Shepard: comment on the fact that $W_k(x) \to \infty$ as $x \to x^{(k)}$. Add a plot showing the shape of $W_k(x)$ for one or two points.
9. page 6, Linear Shepard: what is the definition of W_k in this case?
10. page 6, Box Splines. You claim that this method uses sparse grids and you give reference [33] for details, but I checked it and there is no mention of sparse grids there, where by sparse grids I mean the specific construction that can be found e.g. in [1] as already mentioned. Please remove this sentence.
11. page 7, Voronoi. Add plot to show a Voronoi tassellation, the shape of the function $v^{x^{(i)}}(y)$ and its support.
12. page 7, section 4. You haven’t replied to my previous question: how would you measure the error if the range of approximation is the space of continuous functions (not specifically CDF, i.e. function that can take values outside $[0, 1]$)?
13. page 7: “However, for a CDF [...] the properties are”. Replace “however” with “instead”?

14. page 8, caption of Figure 1: move the sentence "For this example the KS test null hypothesis is rejected at p-value 0.01, however it is not rejected at p-value 0.001" from the caption to the main text, after having introduced the KS formula. I think that the Appendix provided on the meaning of confidence intervals and null hypothesis is too short for someone without a background in statistics.

15. page 8: "that two CDFs come from the same underlying distribution". Maybe change to "empirical CDF"?

16. page 8: I don't understand the expression "round trip prediction methodology".

17. page 10, Lemma 3: add a picture of $g(t), g'(t)$ and of the line w. Give details on these equalities:

$$\frac{g'(0)\tilde{t}}{2} = \int_0^{\tilde{t}} (g'(0) - \gamma_g t)dt$$
$$\frac{-g'(0)^2}{2\gamma_g} = g'(0) - \frac{-g'(0)^2}{2\gamma_g} - \frac{\gamma_g}{2}$$

I was able to derive the first one but not the second one.

18. page 11: "Only linear convergence to the true function can be achieved in practice, without the incorporation of additional observations." I think it has more to see with the spreading of the points. The test on the oscillatory function was done with Fekete points; you should also try with other space filling points, such as Quasi Monte Carlo points, [5]. Also, please add in Figure 1 the reference slopes that show linear and quadratic decay of the error with respect to the number of points. Plot also the Fekete points used in the case $d = 2$ (e.g. the case with 128 or 512 points).

19. page 13, test $d = 20$: could it be that you just need more points before you can see convergence? Or that the Fekete points in $d = 20$ are just bad and you should use Quasi Monte Carlo?

20. page 13, test with $d = 2$:

    - could it be that the MLP stagnation is due to not having done enough epochs of training?
    - What do you need the polynomials of order $n + 1$ for (the Fekete points are obtained by the polynomials of order n)?
    - The sentence "For each test two experiments are executed, one with exact function evaluations (for interpolation) and one with a constant signal-to-noise (SNR) ratio of 10:1 (for regression)" sounds weird: don't you perform both interpolation and regression on both the data sets with and without noise?
    - The sentence "The bound from the theorem suggests that exponentially increasing the number of data points should result in a decreasing error." is imprecise: increasing even exponentially won't help if you place the points badly (e.g. just in one half of the domain). I guess you mean here that using Fekete points and tensorizing them you obtain well-spaced points that guarantee convergence, at the expenses of an exponentially increasing number of points.

21. page 15, test on forest fire: you should mention that timings and lowest absolute prediction error can be checked in the Appendix (same for the other tests). I would add at least for this test a scatter plot of predicted values versus correct values, at least for the two/three best methods (SVR, MLP, Delaunay) for a couple of batches of the k-fold cross validation, to see qualitatively how good is the prediction for small/average/large values of predicted areas. You could already mention here the computational costs. The issue of computational costs is actually very relevant and should be mentioned again explicitly in the conclusions.

References

[1] H.J Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[2] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric pdes. *Foundations of Computational Mathematics*, 14(4):601–633, 2014.

[3] G. Migliorati, F. Nobile, E. Von Schwerin, and R. Tempone. Approximation of quantities of interest in stochastic pdes by the random discrete L^2 projection on polynomial spaces. *SIAM Journal on Scientific Computing*, 35(3):A1440–A1460, 2013.

[4] F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse-grid approximation of Hilbert-space-valued functions: application to random elliptic PDEs. *Numerische Mathematik*, 134(2):343–388, 2016.

[5] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high-dimensional integrals? *J. Complexity*, 14(1):1–33, 1998.