Tuning the interpolation basis in a multigrid decomposition for local error control

Nicolas Vidal, Qian Gong, Viktor Reshniak, Scott Klasky

1 Introduction

1.1 Context

Large scale scientific applications generate an ever-increasing amount of data. The appetite with data in scientific computing is further increased by the development of training models and AI. In comparison, the bandwidth on existing platform is limited. To address these challenges, new methods for compression have to be designed and have been a dynamic field of study.

In scientific workflows, the amount of repeating pattern in the bit representation of the data, due to the potentially random components of the mantissa of the prevalent floating-point representation.

The solution is error-bounded compression.

Additionally, a hierarchical multi-grid decomposition ensures that the most critical portion of the data is transmitted first. This permits users to retrieve the adequate amount of data to achieve the required accuracy after recomposition.

Most of this paper uses MGARD: the Multigrid Adaptive Reduction for data [**ainsworth2018multilevel**, **ainsworth2019multilevel**, **ainsworth20quantitative**] but we have the ambition to provide insight for other related tools.

1.2 Motivation

1.3 Related work

2 Overview of MGARD

- Presentation of compression method
- Presentation of the basis
- Theoretical error bounds for high order
- Literature review

This section serves as a presentation of MGARD, what are its objectives, how it works.

MGARD is a multigrid decomposition-based compressor. The decomposition transform data into coefficients being more suitable to compression. An input of a d-dimensional array u of value u as input is interpreted by MGARD as a continuous function having values u on a grid $\mathcal{N}_L$ having the same structure as the array. We can defined a serie of smaller grids $\mathcal{N}_{L-1}$,...,$\mathcal{N}_0$ downsizing the size by approximately a factor 2^d between two consecutive grids. Each grid being a sample of the previous one. MGARD decomposition starts from level L on the finest grid $\mathcal{N}_L$ and stops at level) on the coarsest, $\mathcal{N}_0$ The decomposition use two operations: L^2 projection, noted Q_l for level land multilinear interpolation, noted Π_l for level l. The multilevel coefficients on $\mathcal{N}_l$ $\mathcal{N}_{l-1}$ are the projectoin $Q_l u$ to the current grid minus its interpolation $\Pi_{l-1} Q_l u$ on the next coarser grid $\mathcal{N}_{l-1}$. $\Pi_{l-1} Q_l u$ is transformed to $Q_{l-1} u$ using the projection of the coefficients to $\mathcal{N}_{l-1}$. The procedure is then repeated at the next level.

We can resume the MGARD decomposition procedure as the follows:

1. The decomposition starts with $Q_L u = u$
2. Compute the piecewise linear interpolant $\Pi_{l-1} Q_l u$ and substract it from $Q_l u$ to get the multilevel coefficients $u_m c$ at level l. These coefficients encode $(I - \Pi_{l-1}) Q_l u$
3. Project the multilevel coefficients to the coarser level to obtain and add the obtained correction to the interpolant $\Pi_{l-1} Q_l u$ to obtain the L^2 projection to the next coarser level, $Q_{l-1} u$.
4. Repeat the above process until l=0.

This process transform the original input into multilevel coefficients u_mc. In our experimental pipeline, we use a simple version of the quantization:

$$\sum_{l=0}^{L} \sum_{x \in \mathcal{N}_l *} |u\_mx[x] - \tilde{u}\_mx[x]| \leq \tau$$

where $\tilde{u}$ is a reduced representation of u after quantization. MGARD provides guaranties on the error $||u - \tilde{u}||$ for L^2 and L^∞.

3 Impact of interpolation basis order

1. Observations on error propagation/local control
2. Impact of the dataset: study the relation between dataset, basis and compression ratio
3. Illustration with small test cases

In this section, we discuss how interpolation basis of different order can have provide different compression ratioi for the same error bound, depending on the data refactored.

3.1 Motivation

A simplified explaination on how different interpolation basis yields different properties follows:

1. The lower the order, the less points it needs
2. Lower order have low error decay and low propagation
3. Higher order provide a smoother approach with fast error decay but more propagation

The accuracy of a decomposition scheme is dependent on the data structure

1. In sparse data set, most values can consist of noise and be close to zero, having a huge impact of the relative error measurement. Sharp variations on the data can also make the interpolation imprecise.
2. In smooth data set, an high level interpolation is close to the data.

The intuition would dictate to use high order interpolation for smooth data and lower order interpolation for sparse data. We explore these relations and illustrate with small test cases.

3.2 Compression pipeline

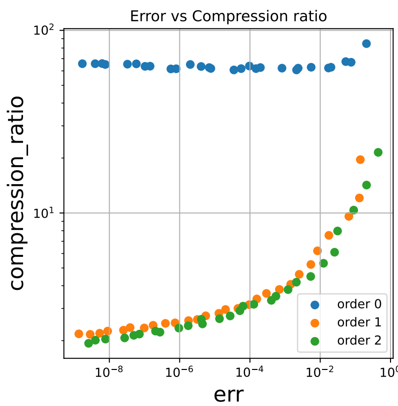
In this section, we summarize the compression procedure used by our adaptative method. TODO: figure. The first two steps represent the original MGARD compression pipeline.

1. Compute MGARD multilevel decomposition using the selected order decomposition base.
2. Error-bounded quantization
3. Huffman encoding of the resulting coefficients
4. Compression using zstd

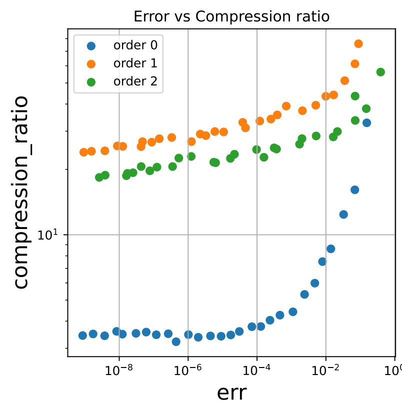
We then measure the relative errors and compression ratio by comparison with the original data.

4 Adaptive solution

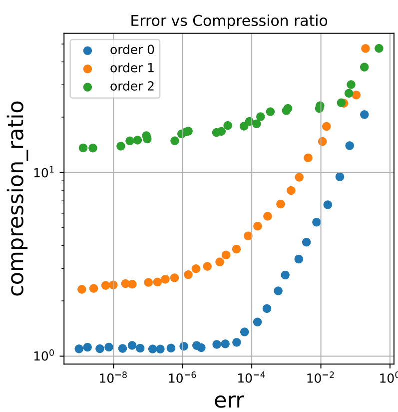
1. Definition of metrics used to pick the basis →residuals
2. Example of picking the correct basis
3. Adaptation over space
   (a) Region partitioning



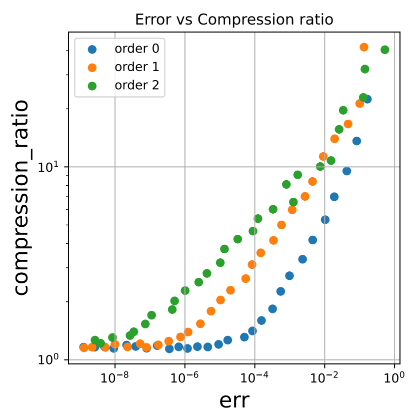
(a) Piece-wise constant data



(b) Piece-wise linear data



(c) Polynomial data (order 2)



(d) Polynomial data (order 4)

Figure 1: Compression ratio obtained for different error depending on the interpolation basis.

(b) Different interpolation for data/background (buffer zone, hard to bound error)

4. Adaptation over time

4.1 Metrics to select the order of basis function

Interpolation error on the finest grid In the MGARD decomposition pipeline, the multilevel coefficients of the coarse nodes of the grid at level l represent the interpolation error at the given level. We use these multilevel coefficient at finest level to evaluate which order is better. Lower coefficients for a given order mean that the underlying interpolation base is to be favored. This can be costly for a large dataset, in this case, a smaller sample can be used. However, the choice of the interpolation order depends on the space-wise distribution of the data and requires a metric computed on the whole region considered.

4.2 Majority voting vs L2

In order to compare two sets of multilevel coefficients, we propose two methods.

1. L2-norm, the order constructing multilevel coefficient with the smallest L2-norm is used for the decomposition.

2. Majority voting: For each datapoint, vote for the order providing the smallest multilevel coefficient. After consideration of the whole sample, the order with the most votes is used for the compression.

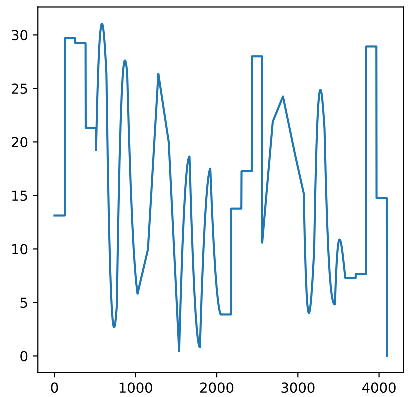
4.3 Minimal example of an adaptive solution

To demonstrate the relevance of adaptive method, we design a simple example. We construct a input function with the following pipeline. Defining a set of n intervals $I_{0\leq i\leq n}$ representing a partition of the definition space, and n orders, we define for each interval I_i a subfunction of order o_i by interpolating between random points (including boundaries points for continuity) with the appropriate langrange polynomial.

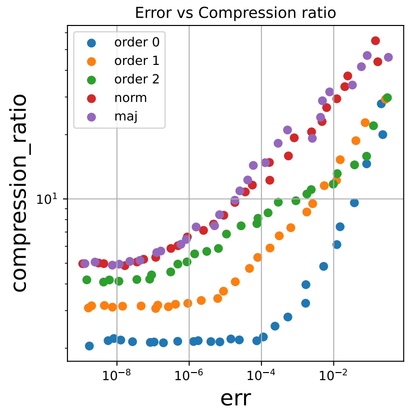
a function is constructed by interpolation using different orders for d

4.4 Interpolation alongside axis

In practice, the MGARD decomposition steps of interpolation and projection are performed one dimension at a time alongside each axis. Each of these step downsampling the size of the grid by a factor 2. In multidimensional scientific data, one dimension often represents time. The variation of the data along this dimension has a specific behavior as compared to the others. In this sense, it can be be relevant to use different order basis for different dimensions.



(a) Example synthetic function with different orders components



(b) Compression ratio compared to reconstruction error for different interpolation order

5 Experimental evaluation & Case study

- Synthesized data
- Real scientific data
- Data which can be partitioned into regions fitting with varied basis functions
- Overhead on memory/computation

6 Conclusion