Research Statement • Ben Whitney

Abstract

My research primarily concerns data compression, specifically multilevel compression methods. The ballooning scale of scientific simulations motivates the use of compression in scientific workflows, and the heterogeneous storage media of extant petascale and planned exascale machines are well-matched to multilevel methods, which split their input into components of varying priorities to be stored on devices of varying capacities and speeds. My work in this field began with an analysis of a lossless compression method based on decimation, a commonly used lossy technique [AKW17]. We derived bounds and computed expectations for compression ratios achieved with deterministic and probabilistic data, respectively. Subsequently, we began work on a suite of original multilevel compression methods called `MGARD` (MultiGrid Adaptive Reduction of Data). `MGARD` is designed to compress scientific data while preserving the features of interest to application scientists. Our initial efforts focused on compression with control of error as measured in L^2 [Ain+18c] and L^∞ [Ain+18a], and a third paper [Ain+18b] added the ability to control the effects of compression on quantities of interest relevant to the data. We paid special attention to the efficient implementation of the method, providing optimal complexity algorithms of the central decomposition and recomposition procedures, and demonstrated its feasibility with applications to a variety of real world datasets. Recent work [Ain+18d] has focused on the case of unstructured grids and on incorporating the numerical methods used to generate simulation data into the compression process. I also have one paper, [Cho+18], in algebraic geometry, which was the result of a project begun while I was an undergraduate and completed during the final year of my graduate studies.

Multilevel Splitting

`MGARD` is applicable to data u given on a nested hierarchy of grids. Such hierarchies are frequently encountered in practice – for instance, when multigrid methods or adaptive mesh resolution is used – and can also be introduced at the compression stage by means of an interpolation from the simulation grid to the compression grid. Let $\mathcal{P}_0, \dots, \mathcal{P}_L$ be the levels in such a hierarchy, with $V_0, \dots, V_L$ associated function spaces, assumed to be piecewise linear finite element spaces. At the heart of `MGARD` is the *multilevel splitting*

$$V_L \ni u \longleftrightarrow ((I - \Pi_{-1})Q_0 u, \dots, (I - \Pi_{L-1})Q_L u) \in V_0 \times \cdots \times V_L \qquad (*)$$

where $\Pi_{\ell-1}$ is the piecewise linear interpolant onto $V_{\ell-1}$ (with Π_{-1} defined to be zero) and Q_ℓ is the L^2 projection onto V_ℓ. The splitting $(*)$ combines the ease of representation of the hierarchical decomposition (with components $(\Pi_\ell - \Pi_{\ell-1})u$) and the superior stability properties of the orthogonal decomposition (with components $(Q_\ell - Q_{\ell-1})u$). It is in fact closely tied to the orthogonal splitting, with efficient $O(N)$ algorithms available for obtaining $(Q_\ell - Q_{\ell-1})u$ from $(I - \Pi_{\ell-1})Q_\ell u$ and *vice versa*. This leads to optimal complexity algorithms for reconstructing $Q_\ell u$ from the first $\ell + 1$ multilevel components $(I - \Pi_{-1})Q_0 u, \dots, (I - \Pi_{\ell-1})Q_\ell u$ and a simple, nonadaptive compression procedure: replace u with its multilevel splitting (an efficient and lossless transformation) and, when a compressed dataset is required, use the first $\ell + 1$ components (with ℓ chosen according to a storage or accuracy constraint) to reconstruct the projection $Q_\ell u$. The connection to the orthogonal decomposition leads to a quasioptimality result [Ain+18c] which establishes that no linear, nonadaptive reduction technique can outperform `MGARD` in the univariate setting by more than a constant factor. This result is extended to the case of general dimension in [Whi18].

Though the methods given in [Ain+18c], being nonadaptive, make limited use of the particulars of the data being reduced, they may still be used to good effect on real datasets. For example, Figure 1 presents the results of compressing the output of a turbulence simulation using nonadaptive `MGARD`.

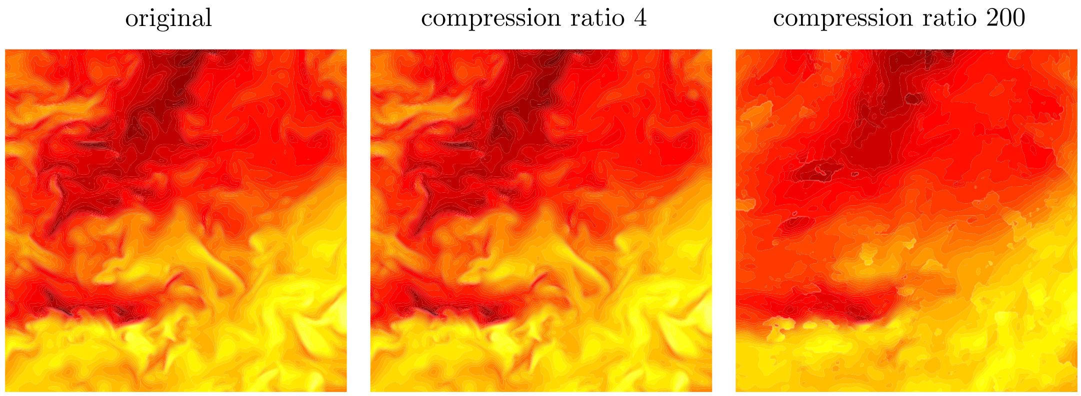


Figure 1: Results of compressing the output of a pseudospectral simulation of forced isotropic turbulence with nonadaptive `MGARD`, taken from [Ain+18c]. The relative errors achieved are 1% for the center plot and 25% for the righthand plot.

The one-dimensional method is applied to the four-dimensional output by fixing points in the spatial domain and gathering the flowfield values observed at those points into timeseries.

Adaptive Reduction

The quasioptimality results and optimal complexity algorithms given in [Ain+18c] suggested that the multilevel splitting $(*)$ could be a fruitful basis for more advanced methods. The following project, [Ain+18a], expands on its predecessor in three aspects: from L^2 to L^∞ error control, from the univariate to the multivariate setting, and from nonadaptive to adaptive reduction. The adaptivity is achieved using the *multilevel coefficients* `u_mc`, a particular representation of the components $(I - \Pi_{\ell-1})Q_\ell u$ of the multilevel splitting. As in [Ain+18c], we give optimal complexity algorithms for transforming the input data u to its multilevel coefficients, with one coefficient `u_mc[`x`]` for each node x of the mesh $\mathcal{P}_L$. Whereas we previously faced a binary choice to retain or discard each multilevel component $(I - \Pi_{\ell-1})Q_\ell u$ in its entirety, we can now make that decision about each multilevel coefficient `u_mc[`x`]` individually. To inform these decisions, we provide reliable and realistic *indicators*, which suggest which of the coefficients contribute most to the components, and *estimators*, which relate changes in the components to changes in the original data u. These functions, and the associated reliability and realism bounds, underlie adaptive reduction algorithms for data given on tensor product grids in multiple dimensions. Figure 2 illustrates an application of these algorithms.

Quantities of Interest

Our most recent paper, [Ain+18b], builds on this work by proving the reliability and realism estimates needed for adaptive reduction in a family of Sobolev norms $\|\cdot\|_r$. More interestingly, we present a technique for compression allowing the preservation of user-specified quantities of interest, like spatiotemporal averages or power spectra. Let $\mathcal{Q}$ be some quantity of interest depending linearly on the data u. If u is to be replaced by some compressed output $\tilde{u}$, we might ask that the compression have limited impact on the quantity of interest: that

$$|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| \leq \tau$$

for some user-prescribed tolerance τ. The ability to produce output respecting such bounds, added without sacrificing the optimal complexity implementation first introduced in [Ain+18c] or the guaranteed error bounds presented in [Ain+18a], is, to the best of our knowledge, unique to `MGARD`. In a demonstration of the flexibility of the technique, we apply it to experimental fusion data while

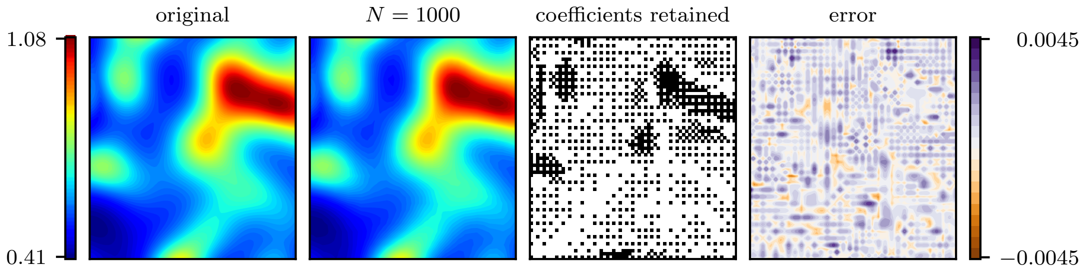


Figure 2: Results of compressing the output of a nonlinear chemical oscillator, the Brusselator, with the adaptive MGARD reduction techniques developed in [Ain+18a]. The original data is represented by 4225 degrees of freedom; the compressed output uses only 1000. The third plot from left indicates the multilevel coefficients retained with black squares. The retained values are clustered in the regions where the data is least regular.

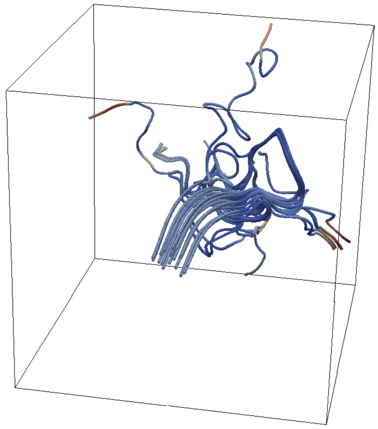
preserving spatial averages to allow 'blob identification' (a procedure of interest to practitioners) to be carried out on the compressed data and to a turbulent combustion simulation while preserving streamlines, an important diagnostic tool similar in spirit to wind tunnel experiments (see Figure 3).

Current Work

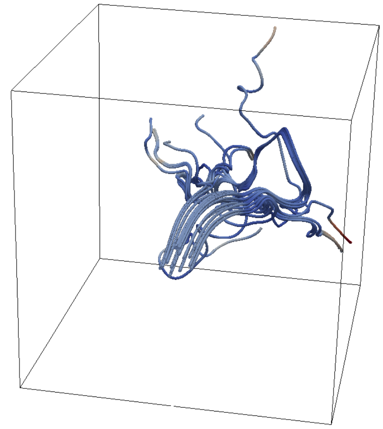
Recent work has focused on expanding MGARD to general mesh hierarchies and incorporating the numerical method used to generate the original data into the compression and reconstruction procedures. Consider, for example, a numerical method outputting a timeseries $u^0, u^1, u^2, \ldots$. MGARD can be applied to each timestep separately, yielding a compressed series $\tilde{u}^0, \tilde{u}^1, \tilde{u}^2, \ldots$, but this approach does not take advantage of any redundancy between the timesteps. One possible remedy, which was investigated in [Ain+18a], is to apply MGARD in time as well as space. This method yields good results as long as multiple timesteps (ideally the entire timeseries) can be loaded into system memory at once. For data too large for this approach, we propose to compute from each compressed timestep $\tilde{u}^i$ an approximation $\overline{u}^{i+1}$ to the following timestep, and to then apply MGARD to the delta $\overline{u}^{i+1} - u^{i+1}$ rather than to u^{i+1} itself. Preliminary results suggest that this technique can improve compression ratios while requiring only moderately more memory and FLOPs than the original strategy of compressing each timestep separately.

Conclusion

My research to date has been born of a marriage between the pursuit for rigorous theoretical bounds and estimates and attention to efficient implementation and real world applicability. I enjoy the interplay between the two sides and find that the theory, in particular, improves the software. I believe that mathematically rigorous compression methods have a basic, foundational place in the future of scientific computing, and I look forward to playing a role in that field's development.



(a) Original dataset.



(b) Reduced dataset with relative error tolerance $\tau = 5 \times 10^{-4}$ and compression ratio 1701.

Figure 3: Results of streamline analysis on the original output of a turbulent combustion simulation (left) and a compressed dataset produced by `MGARD` (right). Careful consideration of the quantity of interest at hand results in streamlines more accurately reproduced than those generated using datasets compressed to similar compression ratios but with compression parameters chosen *a priori* (not pictured; see [Ain+18b]).

References

[AKW17] Mark Ainsworth, Scott Klasky, and Ben Whitney. *Compression Using Lossless Decimation: Analysis and Application.* **SIAM Journal on Scientific Computing** 39 (4), B732–B757. DOI: `10.1137/16M1086248`. 2017-08.

[Ain+18a] Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. *Multilevel Techniques for Compression and Reduction of Scientific Data—The Multivariate Case.* **Submitted**. 2018-01.

[Cho+18] Jed Chou *et al. Diagonal Splittings of Toric Varieties and Unimodularity.* **Proceedings of the American Mathematical Society** 146 (5), pp. 1911–1920. DOI: `10.1090/proc/13902`. 2018-05.

[Whi18] Ben Whitney. *Multilevel Techniques for Compression and Reduction of Scientific Data.* Brown University PhD thesis. URL: `https://repository.library.brown.edu/studio/item/bdr:792867/`. 2018-05.

[Ain+18b] Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. *Multilevel Techniques for Compression and Reduction of Scientific Data—Quantitative Control of Accuracy in Derived Quantities.* **Submitted**. 2018-08.

[Ain+18c] Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. *Multilevel Techniques for Compression and Reduction of Scientific Data—The Univariate Case.* **Computing and Visualization in Science**. DOI: `10.1007/s00791-018-00303-9`. 2018-11.

[Ain+18d] Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. *Multilevel Techniques for Compression and Reduction of Scientific Data—The Unstructured Case.* **In preparation**. 2018-12.