MULTILEVEL TECHNIQUES FOR COMPRESSION AND REDUCTION OF SCIENTIFIC DATA-QUANTITATIVE CONTROL OF ACCURACY IN DERIVED QUANTITIES*

MARK AINSWORTH†, OZAN TUGLUK†, BEN WHITNEY†, AND SCOTT KLASKY‡

Abstract. Although many compression algorithms are focused on preserving pointwise values of the data, application scientists are generally more concerned with derived quantities. Equally well, the user may even be willing to accept a high level of lossiness in the compression provided that the compressed data respect certain invariants, such as mass conservation. In the current work, we develop a mathematical framework and techniques that enable data to be adaptively compressed while maintaining a specified tolerance on a class of user-prescribed quantities. The algorithm is used to augment the functionality of the data reduction package MGARD developed in previous work and the functionality is illustrated by a range of application including data from computational simulation of autocatalytic reaction simulation, turbulent combustion simulation, experimental data obtained from magnetic confinement fusion experiment, and simulation of turbulent flow along a rectangular channel. In each case, we consider one or more relevant quantities of interest and reduce the data so as to preserve these quantities.

1. Introduction. Compression of scientific data is currently highly topical because of the difficulty of transmitting, storing, analyzing, and understanding the drastically increasing quantity of simulation and experimental data being produced. Though compression is typically separated from analysis as a distinct step of the scientific workflow, the two share a common goal: to extract from a mass of raw data the essential structure and key features of the phenomenon under study while ignoring or discarding the noise and simulation errors that have little or no impact on the quantities of interest. So that the compression does not compromise the results of the analysis, it is important to understand how compression methods affect the specific quantities of interest used in the analysis. Lossless methods have unfortunately been generally unable to achieve the high compression ratios needed to handle the huge quantities of data being produced [1, 4]. Consequently, attention has turned to lossy methods, which brings into question the fidelity of the reduced dataset to the original dataset in addition to the amount of compression they achieve. Fidelity and reduction are in direct competition, so it is worthwhile to consider what exactly is required of a reduced dataset in order for it serve as a scientifically useful surrogate.

Consider a typical numerical simulation producing a representation of some field over a spatiotemporal domain. Despite the fact that many compression algorithms are focused on maintaining actual values of the data, application scientists are not

†Division of Applied Mathematics, Brown University, Providence, RI 02912 (Mark_Ainsworth@brown.edu, ozan_tugluk1@brown.edu, ben_whitney@brown.edu).

‡Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 (klasky@ornl.gov).

generally concerned with the pointwise values taken by such a field per se. Rather, the actual quantities of interest are often derived statistics computed from these pointwise values. In principle one could simply compute the relevant quantities of interest in situ and thereby avoid the need to store the field data at all. However, in many applications, specific quantities of interest may become clear only much later in the analysis phase as the postprocessing of the data evolves. Nevertheless, one usually has some a priori notion of a class of quantities of interest which may emerge. By the same token, one may be willing to accept a lossy compression technique provided that the reduced representation maintains certain physical invariants, such as conservation of mass, which would again fall within the above framework.

Suppose that a quantity of interest may be written in the form $\mathcal{Q}(u)$, where u is the function resulting from the numerical simulation and $\mathcal{Q}$ is a bounded linear functional defined on a function space containing u. Let $\tilde{u}$ be some reduced representation of u. As a result of compression, the quantity of interest is calculated with $\tilde{u}$ instead of u, resulting in a loss $\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})$. In preceding work, the authors proposed a reduction procedure that allowed for control of $\|u - \tilde{u}\|_{L^\infty}$ [2]. Although not pursued in [2], we remark that rigorous estimates for loss measured in $L^\infty(\Omega)$ translate into bounds on the loss in the quantity of interest as follows:

$$(1) \qquad |\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| = |\mathcal{Q}(u - \tilde{u})| \leq \Upsilon_{L^\infty}(\mathcal{Q}) \|u - \tilde{u}\|_{L^\infty},$$

where $\Upsilon_{L^\infty}(\mathcal{Q})$ is a constant depending on $\mathcal{Q}$. A disadvantage of this approach is that using a bound on $\|u - \tilde{u}\|_{L^\infty}$ imposes severe restrictions on the reduced representation $\tilde{u}$ at *every point* in the domain, whereas the dependence of the quantity of interest $\mathcal{Q}$ on pointwise values may be rather weak. Ideally, one would use a weaker norm than $L^\infty(\Omega)$ to bound the loss, which should result in greater levels of compression while maintaining a given tolerance on the accuracy of the quantity of interest. In this work we will use a family $\{\|\cdot\|_s : s \in \mathbb{R}\}$ of norms to measure the loss $u - \tilde{u}$. Revisiting (1), we will then bound $|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| \leq \Upsilon_s(\mathcal{Q}) \|u - \tilde{u}\|_s$, where, as before, $\Upsilon_s(\mathcal{Q})$ is a constant depending on $\mathcal{Q}$. More generally, if we are interested in a class of quantities of interest $\{\mathcal{Q}_p(u) : p \in \mathcal{P}\}$, then we can seek to control $\sup_{p \in \mathcal{P}} |\mathcal{Q}_p(u) - \mathcal{Q}_p(\tilde{u})|$ by using the estimate $|\mathcal{Q}_p(u) - \mathcal{Q}_p(\tilde{u})| \leq \Upsilon_s^{\mathcal{P}} \|u - \tilde{u}\|_s$, where $\Upsilon_s^{\mathcal{P}} = \sup_{p \in \mathcal{P}} \Upsilon_s(\mathcal{Q}_p)$. This approach means that any particular quantity of interest $\mathcal{Q}_p$ from the class can be estimated to within the required tolerance.

The remainder of this paper is organized as follows. In the next section we derive estimates for the loss $\|u - \tilde{u}\|_s$ and provide procedures for the computation of $\Upsilon_s(\mathcal{Q})$ and the generation of reduced representations $\tilde{u}$. These are used to define an adaptive algorithm for the reduction of data while maintaining a specifying tolerance in a class of quantities of interest. We demonstrate the flexibility of the approach in the remaining sections, where we consider an autocatalytic reaction simulation, a turbulent combustion simulation, a magnetic confinement fusion experiment, and, finally, a simulation of turbulent flow along a rectangular channel. For each dataset we consider one or more relevant quantities of interest and reduce the data so as to preserve these quantities.

2. A posteriori loss estimator and adaptive reduction. We begin by establishing the setting. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with an initial partition $\mathcal{P}_0$. Let $\mathcal{P}_0$ be repeatedly refined, resulting in a hierarchy $\{\mathcal{P}_\ell : 0 \leq \ell\}$ of meshes on Ω. In practice, refinement is stopped at some maximum level L. Let V_ℓ denote the space of continuous piecewise (multi)linear functions with respect to $\mathcal{P}_\ell$, and let Q_ℓ and Π_ℓ denote the $L^2(\Omega)$ projection and continuous piecewise (multi)linear interpolant,

respectively, onto V_ℓ. Let Q_{-1} denote the zero operator and V_{-1} the trivial function space. Let $\mathcal{N}_\ell$ denote the set of nodes in $\mathcal{P}_\ell$.

Let a function $u \in V_L$ and a bounded linear functional $\mathcal{Q}: V_L \cap L^2(\Omega) \to \mathbb{R}$ be given. Consider the problem of finding a reduced representation $\tilde{u} \in V_L$ (that is, one requiring fewer nonzero degrees of freedom to store than u) such that the loss $|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})|$ is no more than some tolerance τ. Define a family of norms $\{\|\cdot\|_s : s \in \mathbb{R}\}$ on V_L by

$$(2) \qquad \|v\|_s^2 = \sum_{\ell=0}^{L} 2^{2s\ell} \|(Q_\ell - Q_{\ell-1})v\|_0^2$$

with $\|\cdot\|_0$ the $L^2(\Omega)$ norm. Under appropriate conditions, these norms are equivalent [5, 18] to the usual Sobolev norms $\{\|\cdot\|_{H^s} : |s| < \frac{3}{2}\}$. In particular, observe that the s norm $\|\cdot\|_s$ with $s = 0$ is identical to the $L^2(\Omega)$ norm $\|\cdot\|_{L^2}$. As we shall show, adopting the norms $\{\|\cdot\|_s : s \in \mathbb{R}\}$ in place of the $L^\infty(\Omega)$ norm provides additional flexibility which can be used to good advantage. As in (1), the loss in the quantity of interest arising from replacing u by $\tilde{u}$ may be bounded

$$(3) \qquad |\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| = |\mathcal{Q}(u - \tilde{u})| \le \Upsilon_s(\mathcal{Q}) \|u - \tilde{u}\|_s$$

where each $\Upsilon_s(\mathcal{Q}) = \sup_{v \in V_L, v \ne 0} |\mathcal{Q}(v)|/\|v\|_s$ is bounded since V_L has finite dimension. As a result, if $\|u - \tilde{u}\|_s \le \tau/\Upsilon_s(\mathcal{Q})$ for any s, then $|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| \le \tau$, as desired. Two issues must be tackled: first, the computation of $\Upsilon_s(\mathcal{Q})$, and second, the development of a procedure for generating $\tilde{u}$.

2.1. Computation of $\Upsilon_s(\mathcal{Q})$. By the Riesz representation theorem [20] there exists some

$$(4) \qquad \psi_\mathcal{Q} \in V_L \quad \text{such that} \quad \mathcal{Q}(v) = (\psi_\mathcal{Q}, v) \quad \text{for all} \quad v \in V_L,$$

where $(\,\cdot\,, \cdot\,)$ denotes the $L^2(\Omega)$ inner product. For $v \in V_L$, there holds $v = Q_L v = \sum_{\ell=0}^{L}(Q_\ell - Q_{\ell-1})v$ and hence

$$\mathcal{Q}(v) = \sum_{\ell=0}^{L} (\psi_\mathcal{Q}, (Q_\ell - Q_{\ell-1})v).$$

According to the definition of $\{Q_\ell : -1 \le \ell \le L\}$, $(\psi_\mathcal{Q}, w) = ((Q_\ell - Q_{\ell-1})\psi_\mathcal{Q}, w)$ for all $w \in V_\ell \cap V_{\ell-1}^\perp$. In particular, since $(Q_\ell - Q_{\ell-1})v \in V_\ell \cap V_{\ell-1}^\perp$,

$$\mathcal{Q}(v) = \sum_{\ell=0}^{L} ((Q_\ell - Q_{\ell-1})\psi_\mathcal{Q}, (Q_\ell - Q_{\ell-1})v).$$

Applying the Cauchy–Schwarz inequality twice, we find that

$$\mathcal{Q}(v) \le \sum_{\ell=0}^{L} 2^{-s\ell} \|(Q_\ell - Q_{\ell-1})\psi_\mathcal{Q}\|_0 \cdot 2^{s\ell} \|(Q_\ell - Q_{\ell-1})v\|_0 \le \|\psi_\mathcal{Q}\|_{-s} \|v\|_s.$$

Moreover, choosing $v = \sum_{\ell=0}^{L} 2^{-2s\ell}(Q_\ell - Q_{\ell-1})\psi_\mathcal{Q}$, we obtain

$$\mathcal{Q}(v) = \|\psi_\mathcal{Q}\|_{-s}^2 = \|\psi_\mathcal{Q}\|_{-s} \|v\|_s.$$

Consequently,

$$(5) \qquad \Upsilon_s(\mathcal{Q})^2 = \|\psi_\mathcal{Q}\|_{-s}^2 = \sum_{\ell=0}^{L} 2^{-2s\ell} \|(Q_\ell - Q_{\ell-1})\psi_\mathcal{Q}\|_0^2.$$

Introducing the Lagrange basis $\{\phi_\ell(\,\cdot\,;x) : x \in \mathcal{N}_\ell\}$ for V_ℓ, we may write $Q_\ell \psi_{\mathcal{Q}} = \vec{\xi}_\ell^{\mathsf{T}} \vec{\phi}_\ell$, where $\vec{\phi}_\ell$ is the vector with entries $\{\phi_\ell(\,\cdot\,;x) : x \in \mathcal{N}_\ell\}$ and $\vec{\xi}_\ell$ satisfies the linear system

$$M_\ell \vec{\xi}_\ell = \vec{f}_\ell \tag{6}$$

with $M_\ell = \int_\Omega \vec{\phi}_\ell \vec{\phi}_\ell^{\mathsf{T}}$ and $\vec{f}_\ell = \mathcal{Q}(\vec{\phi}_\ell)$. The inclusion $V_{\ell-1} \subseteq V_\ell$ allows us to write $\vec{\phi}_{\ell-1} = R_{\ell-1}\vec{\phi}_\ell$, where $R_{\ell-1}$ is a restriction matrix. As a consequence, using the linearity of $\mathcal{Q}$,

$$\vec{f}_{\ell-1} = \mathcal{Q}(\vec{\phi}_{\ell-1}) = \mathcal{Q}(R_{\ell-1}\vec{\phi}_\ell) = R_{\ell-1}\mathcal{Q}(\vec{\phi}_\ell) = R_{\ell-1}\vec{f}_\ell.$$

The sparsity of $R_{\ell-1}$ means that $\vec{f}_{\ell-1}$ can be computed from $\vec{f}_\ell$ in $O(\dim V_\ell)$ operations. Moreover, the linear system (6) can be solved efficiently, requiring $O(\dim V_\ell)$ operations [2]. Combining these two steps, the hierarchy of projections $\{Q_\ell \psi_{\mathcal{Q}} : 0 \leq \ell \leq L\}$ can be computed recursively from $\vec{f}_L = \mathcal{Q}(\vec{\phi}_L)$ in $O(\dim V_L)$ operations. The functional norm $\Upsilon_s(\mathcal{Q})$ can then be computed using (5). The entire procedure for calculating $\Upsilon_s(\mathcal{Q})$ is summarized in Algorithm 1. Observe that we use the Pythagorean theorem and the definition of $\vec{f}_\ell$ to compute $\|(Q_\ell - Q_{\ell-1})\psi_{\mathcal{Q}}\|_0^2$ in the agglomeration of `square_norm` (line 13 and, similarly, line 15):

$$\begin{aligned}
\|(Q_\ell - Q_{\ell-1})\psi_{\mathcal{Q}}\|_0^2 &= \|Q_\ell \psi_{\mathcal{Q}}\|_0^2 - \|Q_{\ell-1}\psi_{\mathcal{Q}}\|_0^2 \\
&= \vec{\xi}_\ell^{\mathsf{T}} M_\ell \vec{\xi}_\ell - \vec{\xi}_{\ell-1}^{\mathsf{T}} M_{\ell-1} \vec{\xi}_{\ell-1} = \vec{\xi}_\ell^{\mathsf{T}} \vec{f}_\ell - \vec{\xi}_{\ell-1}^{\mathsf{T}} \vec{f}_{\ell-1}.
\end{aligned}$$

Algorithm 1 Calculation of the operator norm of $\mathcal{Q} \colon V_L \to \mathbb{R}$.

1: **function** CALCULATENORM(functional $\mathcal{Q}$, smoothness parameter s)
2:     form mass matrix M_L w.r.t. Lagrange basis $\{\phi_L(\,\cdot\,;x) : x \in \mathcal{N}_L\}$
3:     $\vec{f}_L \leftarrow$ []
4:     **for** i, x in ENUMERATE($\mathcal{N}_L$) **do**
5:         $\vec{f}_L[i] \leftarrow \mathcal{Q}(\phi_L(\,\cdot\,;x))$
6:     **end for**
7:     solve $M_L \vec{\xi}_L = \vec{f}_L$ for $\vec{\xi}_L$
8:     `square_norm` $\leftarrow 0$
9:     **for** $\ell = L-1, \ldots, 0$ **do**
10:         form mass matrix M_ℓ w.r.t. Lagrange basis $\{\phi_\ell(\,\cdot\,;x) : x \in \mathcal{N}_\ell\}$
11:         $\vec{f}_\ell \leftarrow R_\ell \vec{f}_{\ell+1}$
12:         solve $M_\ell \vec{\xi}_\ell = \vec{f}_\ell$ for $\vec{\xi}_\ell$
13:         `square_norm` $\mathrel{+}= 2^{-2s(\ell+1)}(\vec{f}_{\ell+1}^{\mathsf{T}} \vec{\xi}_{\ell+1} - \vec{f}_\ell^{\mathsf{T}} \vec{\xi}_\ell)$
14:     **end for**
15:     `square_norm` $\mathrel{+}= 2^{-2s(0)}(\vec{f}_0^{\mathsf{T}} \vec{\xi}_0)$
16:     **return** $\sqrt{\texttt{square\_norm}}$
17: **end function**

2.2. Adaptive selection of $\tilde{u}$. The first step in developing a procedure for reducing u to $\tilde{u}$ is to decompose u. Define $\Delta_\ell \colon V_L \to V_\ell$ by $\Delta_\ell v = (I - \Pi_{\ell-1})Q_\ell v$ and split u into its images under $\{\Delta_\ell : 0 \leq \ell \leq L\}$:

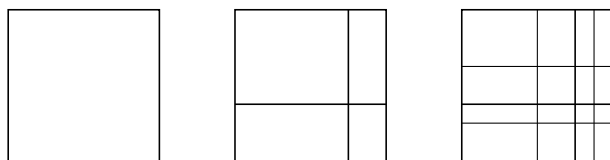$$V_L \ni u \longleftrightarrow (\Delta_0 u, \ldots, \Delta_L u) \in V_0 \times \cdots \times V_L. \tag{7}$$

FIG. 1. *Hierarchy of tensor product grids of the type discussed in Lemmas* 2.1 *and* 2.2 *and Theorem* 2.3. *Note that Lemma* 2.2 *and Theorem* 2.3 *assume* uniform *grids.*

We call $\{\Delta_\ell u : 0 \le \ell \le L\}$ the *multilevel components* of u [2]. The multilevel components do not amount to a proper decomposition of u, as u is not given by the sum of the components. Nevertheless, u may be recovered from the components using the following identity, given in [2]:

$$(8) \qquad \sum_{\ell=0}^{L} (I - Q_{\ell-1})\Delta_\ell u = u.$$

One advantage of using the multilevel components is that

$$\Delta_\ell u = \sum_{x \in \mathcal{N}_\ell^*} \Delta_\ell u(x)\phi_\ell(\,\cdot\,;x),$$

where $\mathcal{N}_\ell^* = \mathcal{N}_\ell \setminus \mathcal{N}_{\ell-1}$. Thanks to (7), this means that u can be completely encoded in a single data structure `u_mc` defined by

$$(9) \qquad \texttt{u\_mc}[x] = \Delta_\ell u(x), \quad x \in \mathcal{N}_\ell^*.$$

We call `u_mc` the *multilevel coefficients* of u. Observe that the multilevel coefficients, like the original representation of u, are represented by $\dim V_L$ scalars. Further details on the multilevel components and coefficients, along with an efficient algorithm for computing them, will be found in [2, 3].

Another advantage of the multilevel components is that the norm $\|v\|_s$ of a function $v \in V_L$ can be estimated in terms of its multilevel components. Define a family of estimators $\{\eta_s \colon V_L \to [0,\infty) : s \in \mathbb{R}\}$ by

$$\eta_s(v)^2 = \sum_{\ell=0}^{L} 2^{2s\ell}\|\Delta_\ell v\|_0^2.$$

The following result shows that $\|\cdot\|_s$ and $\eta_s(\,\cdot\,)$ are equivalent norms.

LEMMA 2.1. *Let $V_0 \subset \cdots \subset V_L$ be spaces of continuous piecewise multilinear functions defined on tensor product grids with mesh ratios $\{\rho_i : 1 \le i \le d\}$ on a domain $\Omega \subset \mathbb{R}^d$. See Figure* 1 *for an illustration. Then*

$$C_d\eta_s(v) \le \|v\|_s \le \eta_s(v) \quad \text{for all} \quad v \in V_L \quad \text{and all} \quad s \in \mathbb{R}$$

with $C_d \ge (\prod_{i=1}^{d} \sqrt{\rho_i} + \sqrt{1/\rho_i})^{-1}$. In particular, $C_d \ge 2^{-d}$ in the case of uniform grids.

Proof. Observe that $(Q_\ell - Q_{\ell-1})\Delta_\ell v = (Q_\ell - Q_{\ell-1})v$ for all $v \in V_L$. As $Q_\ell - Q_{\ell-1}$ is an $L^2(\Omega)$ projection, $\|(Q_\ell - Q_{\ell-1})v\|_0 \le \|\Delta_\ell v\|_0$. The right-hand inequality follows:

$$\|v\|_s^2 = \sum_{\ell=0}^{L} 2^{2s\ell}\|(Q_\ell - Q_{\ell-1})v\|_0^2 \le \sum_{\ell=0}^{L} 2^{2s\ell}\|\Delta_\ell v\|_0^2 = \eta_s(v)^2.$$

For the left-hand inequality, observe that

$$\Delta_\ell v = (I - \Pi_{\ell-1})Q_\ell v = (I - \Pi_{\ell-1})(Q_\ell - Q_{\ell-1})v.$$

The result then follows from Corollary A.3. □

Thanks to Lemma 2.1, (3) now yields

$$|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| = |\mathcal{Q}(u - \tilde{u})| \leq \Upsilon_s(\mathcal{Q})\|u - \tilde{u}\|_s \leq \Upsilon_s(\mathcal{Q})\eta_s(u - \tilde{u}),$$

so that to guarantee that the loss in the quantity of interest is no more than τ it suffices to find $\tilde{u}$ satisfying

$$(10) \qquad \eta_s(u - \tilde{u}) \leq \tau/\Upsilon_s(\mathcal{Q}).$$

Our strategy for generating $\tilde{u}$ is to selectively discard those multilevel coefficients `u_mc`$[x]$ that contribute least to u. We select a subset $\mathcal{N} \subset \mathcal{N}_L$ of coefficients to retain and define `ũ_mc` by

$$\tilde{\texttt{u\_mc}}[x] = \begin{cases} \texttt{u\_mc}[x], & x \in \mathcal{N}, \\ 0 & \text{otherwise.} \end{cases}$$

The reduced representation $\tilde{u}$ is then determined from `ũ_mc` by applying (8) to $\tilde{u}$ and $\Delta_\ell \tilde{u}$ and (9) to $\Delta_\ell \tilde{u}$ and `ũ_mc`. Unfortunately, the impact of zeroing an individual coefficient `u_mc`$[x]$ on the estimator $\eta_s(u - \tilde{u})$ is indirect, making it difficult to gauge the effect of excluding x from $\mathcal{N}$. The next result makes this relation precise.

LEMMA 2.2. *Let $V_0 \subset \cdots \subset V_L$ be spaces of continuous piecewise multilinear functions defined on uniform tensor product grids on a domain $\Omega \subset \mathbb{R}^d$. Let $v \in V_L$, and let* `v_mc` *be the multilevel coefficients for v. Let* $\mathrm{vol}(\mathcal{P}_\ell)$ *denote the (uniform) volume of the elements in $\mathcal{P}_\ell$. Then*

$$C_d \sum_{\ell=0}^{L} 2^{2s\ell} \mathrm{vol}(\mathcal{P}_\ell) \sum_{x \in \mathcal{N}_\ell^*} |\texttt{v\_mc}[x]|^2 \leq \eta_s(v)^2 \leq \sum_{\ell=0}^{L} 2^{2s\ell} \mathrm{vol}(\mathcal{P}_\ell) \sum_{x \in \mathcal{N}_\ell^*} |\texttt{v\_mc}[x]|^2$$

with $C_d \geq 6^{-d}$.

Proof. It suffices to show that $\|\Delta_\ell v\|_0^2 \simeq \mathrm{vol}(\mathcal{P}_\ell)|\texttt{v\_mc}[\mathcal{N}_\ell^*]|^2$ with appropriate constants. Observe that, by the tensor product structure of the mesh, it suffices to show this for $d = 1$. Let $\vec{\alpha} = (\alpha_1, \ldots, \alpha_N)$ be the vector of nodal values of $\Delta_\ell v$ on $\mathcal{N}_\ell$, with α_1 and α_N the values on the endpoints. Using Simpson's rule,

$$\|\Delta_\ell v\|_0^2 = \sum_{i=1}^{N-1} \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \left(\alpha_i^2 + 4(\tfrac{1}{2}\alpha_i + \tfrac{1}{2}\alpha_{i+1})^2 + \alpha_{i+1}^2 \right)$$

$$= \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \sum_{i=1}^{N-1} \begin{bmatrix} \alpha_i \\ \alpha_{i+1} \end{bmatrix}^\mathsf{T} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \alpha_i \\ \alpha_{i+1} \end{bmatrix}.$$

The matrix has eigenvalues 1 and 3, so

$$\|\Delta_\ell v\|_0^2 \geq \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \sum_{i=1}^{N-1} 1 \left| \begin{bmatrix} \alpha_i \\ \alpha_{i+1} \end{bmatrix} \right|^2 = \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \sum_{i=1}^{N}{}'' 2|\alpha_i|^2 \geq \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell)|\vec{\alpha}|^2,$$

$$\|\Delta_\ell v\|_0^2 \leq \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \sum_{i=1}^{N-1} 3 \left| \begin{bmatrix} \alpha_i \\ \alpha_{i+1} \end{bmatrix} \right|^2 = \tfrac{1}{6} \mathrm{vol}(\mathcal{P}_\ell) \sum_{i=1}^{N}{}'' 6|\alpha_i|^2 \leq \mathrm{vol}(\mathcal{P}_\ell)|\vec{\alpha}|^2.$$

Here $\sum''$ denotes the sum with the first and last terms halved. □

Lemma 2.2 means that we can now relate changes in the multilevel coefficients to loss in quantities of interest.

THEOREM 2.3. *Let $V_0 \subset \cdots \subset V_L$ be spaces of continuous piecewise multilinear functions defined on uniform tensor product grids on a domain $\Omega \subset \mathbb{R}^d$. Let $\mathcal{Q}$ be a bounded linear functional on V_L. Let $u \in V_L$ with multilevel coefficients* `u_mc`. *Let* `ũ_mc` *be a set of multilevel coefficients and let $\tilde{u} \in V_L$ be the corresponding function. Then the loss in the quantity of interest is bounded by*

$$(11) \qquad |\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})| \leq \Upsilon_s(\mathcal{Q}) \left(\sum_{\ell=0}^{L} 2^{2s\ell} \operatorname{vol}(\mathcal{P}_\ell) \sum_{x \in \mathcal{N}_\ell^*} |u\_mc[x] - ũ\_mc[x]|^2 \right)^{1/2}.$$

Proof. Define a set `v_mc` of multilevel coefficients by `v_mc[x] = u_mc[x] − ũ_mc[x]` for $x \in \mathcal{N}_L$. By linearity, the corresponding function $v \in V_L$ is equal to $u - \tilde{u}$. The result then follows from (3) and Lemmas 2.1 and 2.2. $\square$

Equation (11) shows that the square of the loss in the quantity of interest incurred by discarding the multilevel coefficient `u_mc[x]`, $x \in \mathcal{N}_\ell^*$, or equivalently choosing `ũ_mc[x] = 0`, is bounded by $2^{2s\ell} \operatorname{vol}(\mathcal{P}_\ell)|u\_mc[x]|^2 \Upsilon_s(\mathcal{Q})^2$. In Algorithm 2 we seek to discard as many coefficients as possible while respecting the bound (10). This can be achieved by a simple modification of the well-known "thresholding" approach advocated in the wavelet literature [10].

Previous work [2, 3] used the multilevel components to define a procedure, MGARD, for the reduction of data while maintaining [3] a bound on $\|u - \tilde{u}\|_{L^2}$ or [2] a bound on $\|u - \tilde{u}\|_{L^\infty}$. In the current work, we will use Algorithm 2 to augment the existing functionality of MGARD to permit adaptive reduction of data while maintaining user-prescribed bounds on $|\mathcal{Q}(u) - \mathcal{Q}(\tilde{u})|$.

3. Application to chemical oscillator data. The Brusselator is a nonlinear partial differential equation modeling a chemical oscillator [22] with governing equations

$$u_t = B + u^2 v - (A+1)u + \alpha \nabla^2 u,$$
$$v_t = Au - u^2 v + \alpha \nabla^2 v.$$

The dataset we reduce, which was also used in [2], comprises 1025 simulation timesteps of the concentrations of two chemical species, u and v, on a spatial domain discretized with a 65×65 grid.

3.1. Peak signal-to-noise ratio control. The mean square error (MSE) is widely used as a measure of accuracy [23]. Beyond the MSE itself, practitioners often adopt closely related measures, such as the peak signal-to-noise ratio (PSNR), which can be defined for a general dataset u and reduced representation $\tilde{u}$ by

$$\operatorname{PSNR}(u, \tilde{u}) = 20 \log_{10} \left[\frac{\max(u) - \min(u)}{|\Omega|^{-1/2} \|u - \tilde{u}\|_0} \right].$$

Recall that $\|\cdot\|_0$ is the $L^2(\Omega)$ norm. Lemma 2.1 allows us to bound the PSNR in terms of the estimator $\eta_0(u - \tilde{u})$:

$$\operatorname{PSNR}(u, \tilde{u}) = 20 \log_{10} \big[\max(u) - \min(u)\big] + 10 \log_{10} |\Omega| - 20 \log_{10} \|u - \tilde{u}\|_0$$
$$\geq 20 \log_{10} \big[\max(u) - \min(u)\big] + 10 \log_{10} |\Omega| - 20 \log_{10} \eta_0(u - \tilde{u}).$$

Algorithm 2 Adaptive reduction to meet a loss tolerance.

---

function REDUCE(multilevel coefficients `u_mc`, tolerance τ, smoothness parameter s, functional norm $\Upsilon_s(\mathcal{Q})$)

    procedure INDICATOR(node x)

        find $\ell \in \{0, \ldots, L\}$ so that $x \in \mathcal{N}_\ell^*$

        return $2^{2s\ell}\operatorname{vol}(\mathcal{P}_\ell)\,|\texttt{u\_mc}[x]|^2\,\Upsilon_s(\mathcal{Q})^2$

    end procedure

    procedure MULTILEVELCOEFFICIENTS(cutoff C)

        $\tilde{\texttt{u}}\texttt{\_mc} \leftarrow \texttt{[]}$

        for $x \in \mathcal{N}_L$ **do**

            $\tilde{\texttt{u}}\texttt{\_mc}[x] \leftarrow \texttt{u\_mc}[x]$ **if** INDICATOR(x) $> C$ **else** 0

        end for

        return $\tilde{\texttt{u}}\texttt{\_mc}$

    end procedure

    $C_{\mathrm{hi}} \leftarrow \max_{x \in \mathcal{N}_L}$ INDICATOR(x) $+ 1$

    $C_{\mathrm{lo}} \leftarrow \min_{x \in \mathcal{N}_L}$ INDICATOR(x) $- 1$

    //We need to check whether C_{hi} is in fact too high. If τ is very large,

    //we may be able to discard all of the multilevel coefficients.

    $\tilde{\texttt{u}}\texttt{\_mc} \leftarrow$ MULTILEVELCOEFFICIENTS(C_{hi})

    if $\eta_s(u - \tilde{u}) \le \tau/\Upsilon_s(\mathcal{Q})$ **then return** $\tilde{\texttt{u}}\texttt{\_mc}$

    while $C_{\mathrm{lo}} <$ INDICATOR(x) $\le C_{\mathrm{hi}}$ for any $x \in \mathcal{N}_L$ **do**

        $C \leftarrow (C_{\mathrm{hi}} + C_{\mathrm{lo}})/2$

        $\tilde{\texttt{u}}\texttt{\_mc} \leftarrow$ MULTILEVELCOEFFICIENTS(C)

        if $\eta_s(u - \tilde{u}) > \tau/\Upsilon_s(\mathcal{Q})$ **then** $C_{\mathrm{hi}} \leftarrow C$ **else** $C_{\mathrm{lo}} \leftarrow C$

    end while

    return MULTILEVELCOEFFICIENTS(C_{lo})

end function

So, choosing $\tilde{u}$ to satisfy

$$\eta_0(u - \tilde{u}) \le 10^{-\tau/20}|\Omega|^{1/2}\big[\max(u) - \min(u)\big]$$

will guarantee that the PSNR is at least τ. Figure 2 shows the results of reducing a single timestep of the Brusselator simulation using `MGARD` with a range of prescribed tolerances on the PSNR. As expected, the PSNR achieved by the reduced datasets always meets the prescribed tolerance and, moreover, is close to the prescribed value.

3.2. Pixel value control. In practice, users may be less concerned with PSNR than with a more tangible measure of quality, such as the fidelity of a visualization produced using the reduced representation. Given a dataset u to be visualized, we might ask that the reduced representation $\tilde{u}$ satisfy the condition that each pixel of the visualization produced with $\tilde{u}$ be within some tolerance τ of the corresponding pixel of the visualization produced with u. More precisely, assume that each pixel $p \in \mathcal{P}$ of the visualization corresponds to some rectangular subset p of the domain Ω and that the value of that pixel is given by the average $\mathcal{Q}_p(u)$ of u over p:

$$\mathcal{Q}_p(u) = \fint_p u(x)\,\mathrm{d}x.$$

The pixel fidelity condition then becomes

(12) $$\qquad |\mathcal{Q}_p(u) - \mathcal{Q}_p(\tilde{u})| \le \tau \quad \text{for all} \quad p \in \mathcal{P},$$
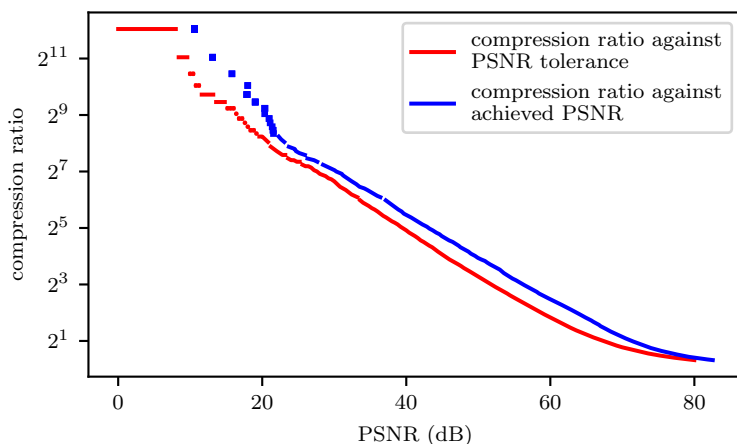
FIG. 2. *Illustration of results obtained when reducing timestep* 300 *of the Brusselator simulation to a range of PSNR tolerances.*

which falls within the general framework described in section 2, wherein we seek to reduce the data while controlling the loss in some quantity of interest rather than the PSNR.

Our framework is based on utilizing bounds of the form

$$|\mathcal{Q}_p(u) - \mathcal{Q}_p(\tilde{u})| \leq \Upsilon_s(\mathcal{Q}_p)\|u - \tilde{u}\|_s \quad \text{with} \quad s \in \mathbb{R}.$$

How to choose s? We wish to select a value that results in MGARD discarding as many multilevel coefficients u_mc[x] as possible. This is achieved by choosing s so as to minimize the multiplicative factor $2^{2s\ell}\Upsilon_s(\mathcal{Q}_p)^2$ appearing in Algorithm 2. As s becomes increasing negative, $2^{2s\ell}$ shrinks while $\Upsilon_s(\mathcal{Q}_p)^2$ grows due to the presence of the $2^{-2s\ell}$ factors in (5). In order to gain insight on the best choice of s, we exploit the equivalence of the norms $\|\cdot\|_s$ and the standard Sobolev norms $\|\cdot\|_{H^s(\Omega)}$ when $|s| < \frac{3}{2}$ [5, 18]. With χ_p the characteristic function of p, we can write

$$\mathcal{Q}_p(v) = \fint_p v(x)\,\mathrm{d}x = \frac{1}{|p|}\int_\Omega v(x)\chi_p(x)\,\mathrm{d}x = \frac{1}{|p|}(\chi_p, v).$$

The functional norm $\Upsilon_s(\mathcal{Q}_p)$ is then related to the Sobolev norm $\|\chi_p\|_{H^{-s}}$ as follows:

$$|p|\Upsilon_s(\mathcal{Q}_p) = \sup_{v \in V_L}\frac{(\chi_p, v)}{\|v\|_s} \simeq \sup_{v \in V_L}\frac{(\chi_p, v)}{\|v\|_{H^s}} \leq \sup_{v \in \cup_{\ell \in \mathbb{N}} V_\ell}\frac{(\chi_p, v)}{\|v\|_{H^s}} = \|\chi_p\|_{H^{-s}}$$

for all $|s| < \frac{3}{2}$. Recalling that the characteristic function χ_p belongs to the Sobolev spaces $H^r(\Omega)$ for all $r < \frac{1}{2}$, the above quantity is finite provided $-s < \frac{1}{2}$, or, equally well, $s > -\frac{1}{2}$. Consequently, we choose $s = -\frac{1}{2}$. The exclusion of χ_p from $H^{1/2}(\Omega)$ causes no problems, as the finite dimensionality of V_L guarantees that $\Upsilon_{1/2}(\mathcal{Q}_p) < \infty$.

We apply MGARD to timestep 480 of the Brusselator simulation with smoothness parameter $s = -\frac{1}{2}$ and tolerance $\tau = 5 \times 10^{-2}$. The functional norm used is $\max_{p \in \mathcal{P}} \Upsilon_{1/2}(\mathcal{Q}_p)$, so that (12) is satisfied. Figure 3 displays the visualizations of the original and reduced datasets, indicating in the lower left which of the multilevel coefficients are retained. Observe that the loss at every pixel in the visualization meets the specified tolerance, as guaranteed by the results of section 2.
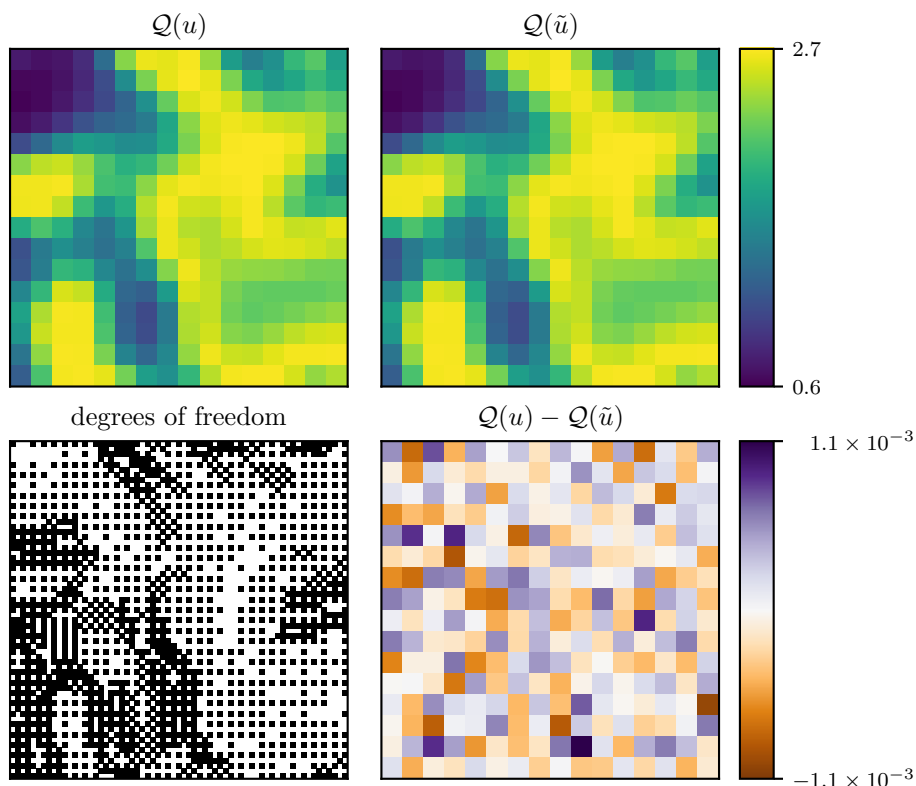
FIG. 3. *Illustration of results obtained when reducing timestep* 480 *of the Brusselator simulation using pixels of size* 4×4 *with a tolerance of* $\tau = 5 \times 10^{-2}$.

3.3. Adaptive decimation. A similar need for reduction while preserving averages occurs in the context of timestepping applications. In order to maintain stability, numerical methods for PDEs often use a simulation timestep significantly smaller than the timestep ultimately needed to resolve the phenomenon of interest, resulting in an excessively large output. To fix ideas, suppose a simulation is carried out over a spatial domain Ω and a temporal domain $[0, T]$, with the latter subdivided into a collection $\mathcal{P}$ of time intervals. In practice, data reduction is often achieved using *decimation* [1], whereby each time series $\{u(x, t) : t \in p\}$, $x \in \Omega$, $p \in \mathcal{P}$, is replaced by either a single datum $u(x, t_*)$ for some $t_* \in p$ or (preferably) the average $\mathcal{Q}_p(u(x, \cdot))$ of the values in p, given by, with χ_p the characteristic function of p,

$$\mathcal{Q}_p(u(x, \cdot)) = \fint_p u(x, t)\, \mathrm{d}t = \frac{1}{|p|} \fint_{[0,T]} u(x, t) \chi_p(t)\, \mathrm{d}t.$$

Such indiscriminate data reduction can lead to, on the one hand, undersampling in regions of high variability and, on the other, oversampling in regions of low variability.

Instead, we propose to adaptively generate a reduced representation $\tilde{u}(x, \cdot)$ of the original time series $u(x, \cdot)$ satisfying, for some user-prescribed tolerance τ,

$$|\mathcal{Q}_p(u(x, \cdot)) - \mathcal{Q}_p(\tilde{u}(x, \cdot))| \leq \tau \quad \text{for all} \quad p \in \mathcal{P}.$$

We take the dataset to be the output of the Brusselator simulation and the partition $\mathcal{P}$ to comprise time intervals formed from 16 consecutive timesteps (i.e., each $p \in \mathcal{P}$

TABLE 1

Compression ratios obtained using a decimation stride of 2^4 and a variety of tolerances and reduction norms. The compression ratios displayed are the harmonic averages over each of the 4225 time series $u(x, \cdot)$ in the Brusselator dataset.

	τ		
	10^{-2}	10^{-4}	10^{-6}
$\|\cdot\|_{-1/2}$	6.4	1.4	1.0
$\|\cdot\|$	4.9	1.2	1.0
$\|\cdot\|_{L^\infty}$	3.1	1.1	1.0

has the form $p = [16i\Delta t, 16(i+1)\Delta t)$. Arguing as above, we use the bound

$$|\mathcal{Q}_p(u(x, \cdot)) - \mathcal{Q}_p(\tilde{u}(x, \cdot)| \leq \Upsilon_s(\mathcal{Q}_p)\|u(x, \cdot) - \tilde{u}(x, \cdot)\|_s$$

and apply Algorithm 2 in MGARD with the choice $s = -\frac{1}{2}$. We repeat this procedure with tolerances $\tau \in \{10^{-2}, 10^{-4}, 10^{-6}\}$. For comparison, we also reduce the data while controlling two other measures of loss. First, we use MGARD with smoothness parameter $s = 0$ to reduce the data while controlling the PSNR. Second, we use the variant of MGARD developed in [2] to reduce the data while controlling the pointwise error between $u(x, \cdot)$ and $\tilde{u}(x, \cdot)$. Ensuring that $\|u(x, \cdot) - \tilde{u}(x, \cdot)\|_{L^\infty}$ is no more than $|p|\tau/\|\chi_p\|_{L^1}$ guarantees that the loss in $\mathcal{Q}_p$ is no more than τ, as desired, thanks to the inequality

$$|\mathcal{Q}_p(u(x, \cdot)) - \mathcal{Q}_p(\tilde{u}(x, \cdot))| \leq \frac{1}{|p|}\|u(x, \cdot) - \tilde{u}(x, \cdot)\|_{L^\infty([0,T])}\|\chi_p\|_{L^1([0,T])}.$$

The overall compression ratios (averaged over all $x \in \Omega$) for each combination of the three methods and the three tolerances are collected in Table 1. In all cases, the reduced dataset obtained meets the prescribed tolerance τ. The best compression is achieved with $s = -\frac{1}{2}$, which is consistent with the arguments given earlier.

4. Application to turbulent combustion data. In this section we consider data generated by S3D [12], an exascale direct numerical simulation code for turbulent combustion. The dataset in question is from a statistically planar turbulent combustion simulation. The simulation domain is a periodic box discretized with $500 \times 500 \times 500$ spatial gridpoints. The dataset comprises 51 time instances, with a total size of approximately 500 GB. S3D outputs 11 variables (species) at each spatial gridpoint and each timestep of the simulation.

We apply MGARD to each component in this vector-valued dataset separately, since data ranges vary significantly between the species, so that treating each of the species independently avoids discarding finer scale details. Hence the compression cycle for each three-dimensional time instance consists of 11 invocations of MGARD. The compression ratio reported for each experiment is the ratio of the sum of the sizes of the original components and the sum of the sizes of the reduced components.

4.1. Peak signal-to-noise ratio control. We begin by reducing the S3D dataset while controlling the PSNR. The simulation data is reduced with MGARD using a range of target PSNRs. The corresponding compression ratios are compared with those obtained using two state-of-the-art compressors, SZ [7, 21] and ZFP [8, 15]. Both MGARD and SZ feature native PSNR error control, meaning that the desired PSNR can be set beforehand and the compression ratio computed from the output. The results

TABLE 2
Compression ratios achieved by MGARD *and* SZ *on the S3D dataset at various PSNR tolerances.*

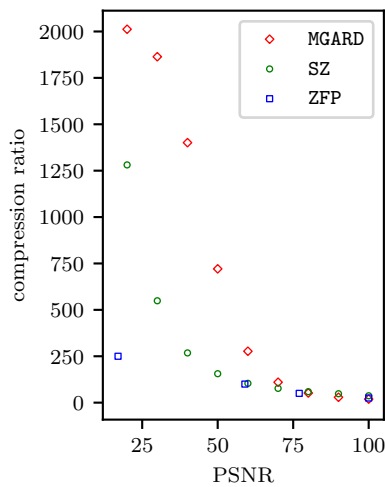| PSNR (dB)	MGARD	SZ
20	2012.5	1281.5
30	1864.8	549.4
40	1401.3	269.0
50	721.1	155.9
60	277.4	103.6
70	110.6	76.6
80	52.2	58.7
90	29.0	48.2
100	18.7	36.9

FIG. 4. *Compression ratios achieved by* MGARD, SZ, *and* ZFP *on the S3D dataset at various PSNR tolerances. At each PSNR, a box-plot of the compression ratios of each species of the output is shown. The target PSNR is set directly for* MGARD *and* SZ, *while for* ZFP *the fixed-rate mode is used and the resulting PSNR is calculated.*
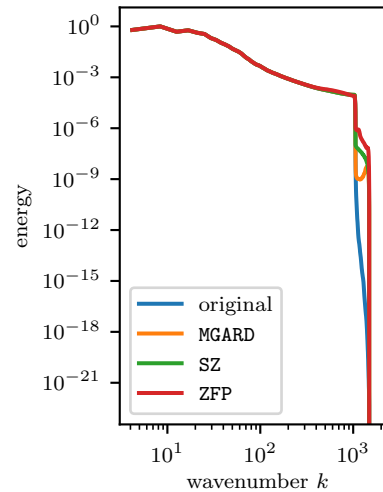
FIG. 5. *Turbulent kinetic energy spectra of the S3D flow field at time 2.999×10^{-3}. The reduced datasets have target PSNRs of 60 dB. The distortion in the spectra is concentrated in the smaller scales of the flow. The larger scales (lower wavenumbers) are faithfully reproduced.*

from these compressors are shown in Table 2. Conversely, ZFP does not offer native PSNR control. Therefore, we call ZFP with a range of prescribed values of the compression ratio and evaluate the PSNR obtained. Figure 4 displays the PSNRs and the compression ratios obtained by all three compressors.

To investigate the effect of PSNR reduction on this dataset, we examine (for a quantitative measure) instantaneous kinetic energy spectra and (for a qualitative measure) instantaneous flame fronts of the original and reduced flows. Figure 5 displays the energy spectra of the original and reduced (to target PSNR 60 dB) flows at the final timestep of the simulation. MGARD, SZ, and ZFP all introduce some distortion in the smaller scales of the flow. The larger scales, meanwhile, are well preserved. Figure 6 displays flame fronts of the original and two reduced (MGARD at 60 dB and MGARD at 40 dB) datasets at the final timestep. Even at low PSNR, the reduced flame front is visually indistinguishable from the original.
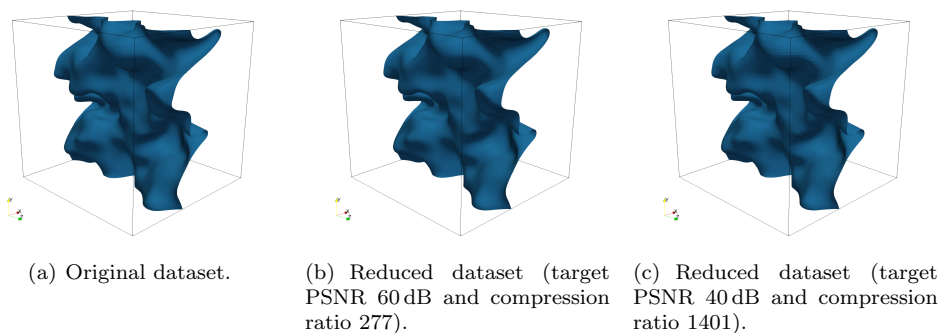
(a) Original dataset.   (b) Reduced dataset (target PSNR 60 dB and compression ratio 277).   (c) Reduced dataset (target PSNR 40 dB and compression ratio 1401).

FIG. 6. *Flame fronts of the original dataset and reduced representations generated by* `MGARD` *at time* 2.999×10^{-3}. *For each dataset the contour of the flow at instantaneous temperature* $1200\,\mathrm{K}$ *is shown.*

4.2. Upscaling and local mass conservation. Using PSNR as a measure of distortion caused by data reduction is commonplace, but PSNR control does not guarantee preservation of physical invariants in the reduced dataset. For the S3D dataset, an important consideration is the local conservation of the mass of the various reactants. The Kolmogorov length scale for the dataset is $1.87\Delta h \sim 2\Delta h$, where Δh is the mesh size [12]. Consequently, it is desirable to enforce local mass conservation on the set $\mathcal{P}$ of cells of size $2\Delta h \times 2\Delta h \times 2\Delta h$, corresponding to conserving the averages over a mesh coarsened by a factor of 2, a common setup in Large Eddy Simulations.

As before, we define for each rectangle $p \in \mathcal{P}$ of the characteristic scale a functional $\mathcal{Q}_p$ by, with χ_p the indicator function of p,

$$\mathcal{Q}_p(u) = \fint_{\Omega} u(x)\,\chi_p(x)\,\mathrm{d}x.$$

Following subsections 3.2 and 3.3, this approach suggests using `MGARD` with smoothness parameter $s = -\frac{1}{2}$. We reduce the data so that the mass fractions over the cells are within a tolerance of $\tau = 10^{-5}$ of the true values.

Figures 7 and 8 display the mass fraction contours of the original dataset and datasets reduced so as to preserve averages over cells of size $2\Delta h \times 2\Delta h \times 2\Delta h$ and $10\Delta h \times 10\Delta h \times 10\Delta h$, respectively. The only difference between the two reductions are the values taken by the norms $\Upsilon_{1/2}(\mathcal{Q}_p)$. Though the cubes of side length $2\Delta h$ are smaller than the cubes of side length $10\Delta h$, the norms of the associated functionals are actually larger. As a result, the scaled tolerances $\tau/\Upsilon_{1/2}(\mathcal{Q}_p)$ are smaller. This is not entirely unexpected: the condition that averages on a certain scale be preserved becomes more stringent as the scale shrinks.

4.3. Streamline control. An important tool in the computational scientist's arsenal is *streamlines*, which are computed in a process akin to wind tunnel experiments. Virtual dye is injected into a flow field u and the resulting traces help visualize the flow dynamics. In Figure 9(a), streamlines generated using the original S3D dataset and a variety of initial conditions are plotted. In general, streamlines generated using a reduced representation $\tilde{u}$ will differ from those generated using u. Figure 9(b) displays streamlines produced using a reduced flow field generated by `MGARD` with target PSNR 40 dB (compression ratio 1403). Though the streamlines are similar to the true trajectories in Figure 9(a), there is some noticeable divergence.
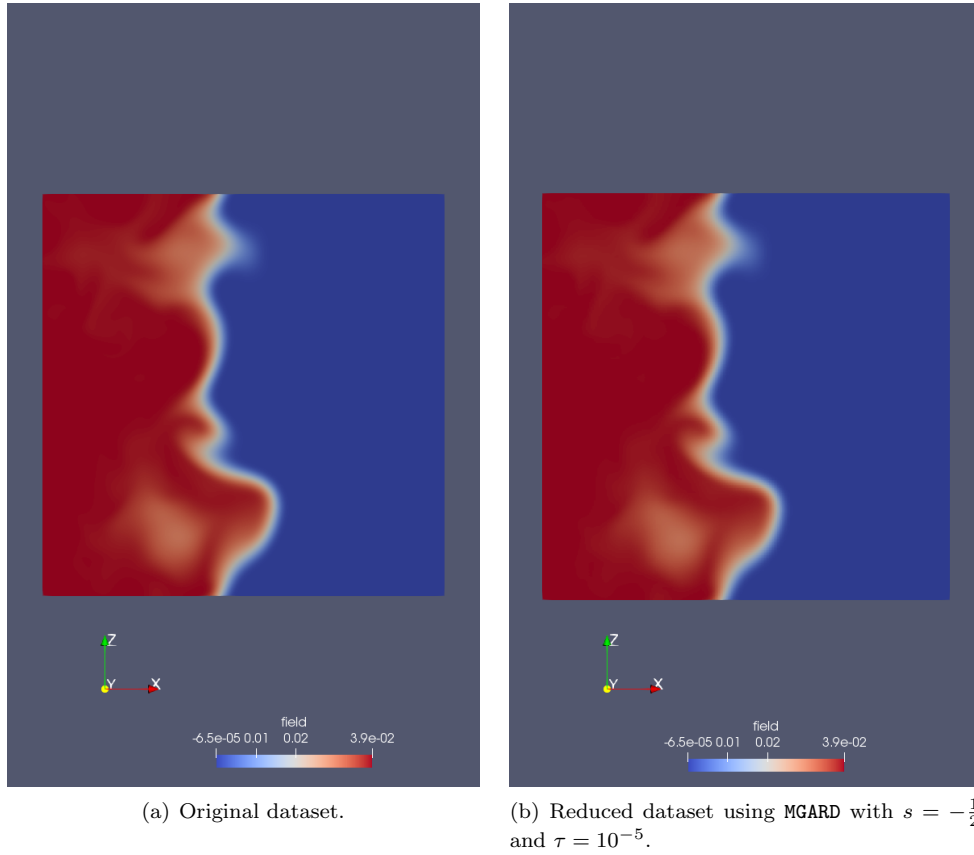
(a) Original dataset.

(b) Reduced dataset using `MGARD` with $s = -\frac{1}{2}$ and $\tau = 10^{-5}$.

FIG. 7. *Contours of fuel concentration in a plane orthogonal to the y axis. The data are reduced so that mass fractions over cells of size $2\Delta h \times 2\Delta h \times 2\Delta h$ are preserved to within a tolerance of $\tau = 10^{-5}$.*

Can we produce better streamlines while maintaining the same level of compression by controlling an appropriate functional and making a more judicious choice of s (compared with the choice $s = 0$ when controlling PSNR)?

Let $\Omega \subset \mathbb{R}^d$ be a spatial domain on which a velocity field $u \colon \Omega \to \mathbb{R}^d$ is defined. Given an initial starting point $x_0 \in \Omega$, a streamline $x(t)$ is defined by the integral equation

$$x(t) = x_0 + \int_0^t u(x(t))\,\mathrm{d}t \quad \text{for} \quad t \in [0, T].$$

An approximate streamline $\tilde{x}(t)$ can similarly be computed from a reduced velocity field $\tilde{u}$ by $\tilde{x}(t) = x_0 + \int_0^t \tilde{u}(\tilde{x}(t))\,\mathrm{d}t$ for $t \in [0, T]$. The error between the streamlines then grows according to

$$(13) \qquad \frac{\mathrm{d}}{\mathrm{d}t}\big[x(t) - \tilde{x}(t)\big] = u(x(t)) - \tilde{u}(\tilde{x}(t)) \quad \text{for} \quad t \in [0, T]$$

$$= \big[u(x(t)) - u(\tilde{x}(t))\big] + \big[u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t))\big]$$

with initial condition $\tilde{x}(0) = x(0) = x_0$. Consider the local error generated in the trajectory by incrementing t from t_k to t_{k+1}. That is, assume that the trajectory

(a) Original dataset.

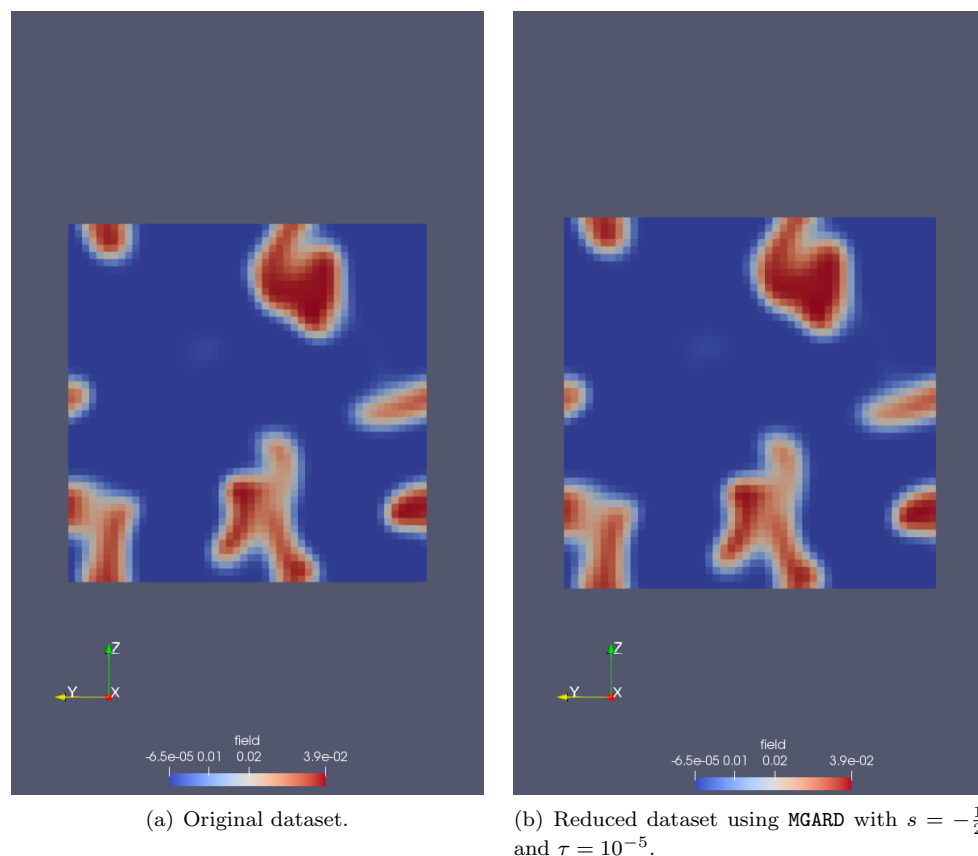(b) Reduced dataset using `MGARD` with $s = -\frac{1}{2}$ and $\tau = 10^{-5}$.

FIG. 8. *Contours of fuel concentration in a plane orthogonal to the x axis. The data are reduced so that mass fractions over cells of size $10\Delta h \times 10\Delta h \times 10\Delta h$ are preserved to within a tolerance of $\tau = 10^{-5}$.*

starts at time $t = t_k$ and consider the problem (13) subject to the initial condition $\tilde{x}(t_k) = x(t_k)$, so that x and $\tilde{x}$ coincide at t_k. If $t_{k+1} - t_k \ll 1$, the linearization error can be neglected and we obtain the equation

$$\frac{d}{dt}\big[x(t) - \tilde{x}(t)\big] - \nabla u(x(t_k)) \cdot \big[x(t) - \tilde{x}(t)\big] = u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t)) \quad \text{for} \quad t \in [t_k, t_{k+1}].$$

This equation shows that the error incurred in the trajectory over $[t_k, t_{k+1}]$ originates from the "source" term $u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t))$. Consequently, by Duhamel's principle, the error incurred in the step is bounded by a multiple of

$$\int_{t_k}^{t_{k+1}} \exp\big[(t_{k+1} - t)\|\nabla u\|_{L^\infty}\big] \big|u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t))\big| \, dt,$$

or, in turn,

$$\exp\big[(t_{k+1} - t_k)\|\nabla u\|_{L^\infty}\big] \int_{t_k}^{t_{k+1}} \big|u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t))\big| \, dt.$$

Based on the right-hand integral, we define a linear functional $\mathcal{Q}_k$ by, for $v \colon \Omega \to \mathbb{R}$

sufficiently smooth,

$$\mathcal{Q}_k(v) = \int_{t_k}^{t_{k+1}} v(\tilde{x}(t))\,\mathrm{d}t.$$

Let χ_k be the characteristic function of the trajectory $\{\tilde{x}(t) : t \in [t_k, t_{k+1}]\}$ of $\tilde{x}$ from t_k to t_{k+1}. Observe that

$$\begin{aligned}
\mathcal{Q}_k(v) &= \int_{t_k}^{t_{k+1}} v(\tilde{x}(t)) \frac{|\tilde{u}(\tilde{x}(t))|}{|\tilde{u}(\tilde{x}(t))|}\,\mathrm{d}t \\
&\leq \frac{1}{m_k} \int_{t_k}^{t_{k+1}} v(\tilde{x}(t))\,|\tilde{u}(\tilde{x}(t))|\,\mathrm{d}t \qquad \text{where} \quad m_k = \min_{t \in [t_k, t_{k+1}]} |\tilde{u}(\tilde{x}(t))| \\
&= \frac{1}{m_k} \int_\Omega v(x)\,\chi_k(x)\,\mathrm{d}x.
\end{aligned}$$

If $t_{k+1} - t_k \ll 1$, then v does not change sign on $[t_k, t_{k+1}]$ and so $\mathcal{Q}_k(|v|) \approx |\mathcal{Q}_k(v)|$. Thus,

$$\int_{t_k}^{t_{k+1}} |u(\tilde{x}(t)) - \tilde{u}(\tilde{x}(t))|\,\mathrm{d}t = \mathcal{Q}_k(|u - \tilde{u}|) \approx |\mathcal{Q}_k(u - \tilde{u})| \leq \frac{1}{m_k} \left| \int_\Omega (u - \tilde{u})(x)\,\chi_k(x)\,\mathrm{d}x \right|.$$

We show in Lemma B.1 that, under certain conditions, $\chi_k \in H^r(\Omega)$ for $r < \frac{1}{2}$ when $d = 1$ and for $r < (1-d)/2$ when $d > 1$. The present simulation is three-dimensional, which suggests that we should control $\|u - \tilde{u}\|_s$ with $s = 1$ to limit the distortion of the streamlines. Figures 9(c) and 9(d) show the results obtained using $s = \frac{1}{2}$ and $s = 1$, respectively. Compared with the choice $s = 0$ (Figure 9(b)), the same order of compression is attained and the streamlines are more faithfully reproduced, with the choice $s = 1$ providing the better results.

5. Application to fusion experimental data. Next we consider data from a magnetic confinement fusion experiment [25]. In the experiment, plasma circulating around a tokamak is kept in place by a magnetic field. This magnetic confinement can be interrupted by spatiotemporal instabilities known as *blobs*. Blobs are regions of increased plasma density and temperature that are aligned to the magnetic field [9]. They can cause heat loss, confinement degradation, and erosion of the containment wall by convecting plasma radially outward [24]. Identifying the blobs is the first step in understanding their dynamics and their interactions with larger-scale flows in the plasma [17].

One of the techniques used to identify blobs is *gas puff imaging* (GPI), wherein a neutral gas puff is introduced into the plasma and the ensuing line emission is imaged [16]. The GPI experiment outputs a sequence of 64×80 pixel images. In this section we apply MGARD to a 200,000-frame selection of that sequence. The data are preprocessed by normalizing about the mean and applying Gaussian denoising over 5×5 pixel regions [25].

5.1. Average control. In this subsection we investigate the effects of MGARD reduction on blob identification. Utilities from the computer vision library openCV [6] are used to identify blob centers and boundaries in both the original frames and frames subjected to an MGARD compression-decompression cycle. In the plasma, blobs exist as regions of increased density and temperature. In the GPI output, they appear as areas of increased brightness [16]. Accordingly, in order to preserve the blobs we wish to limit the loss in local average brightness. As in subsections 3.2 and 4.2, this leads
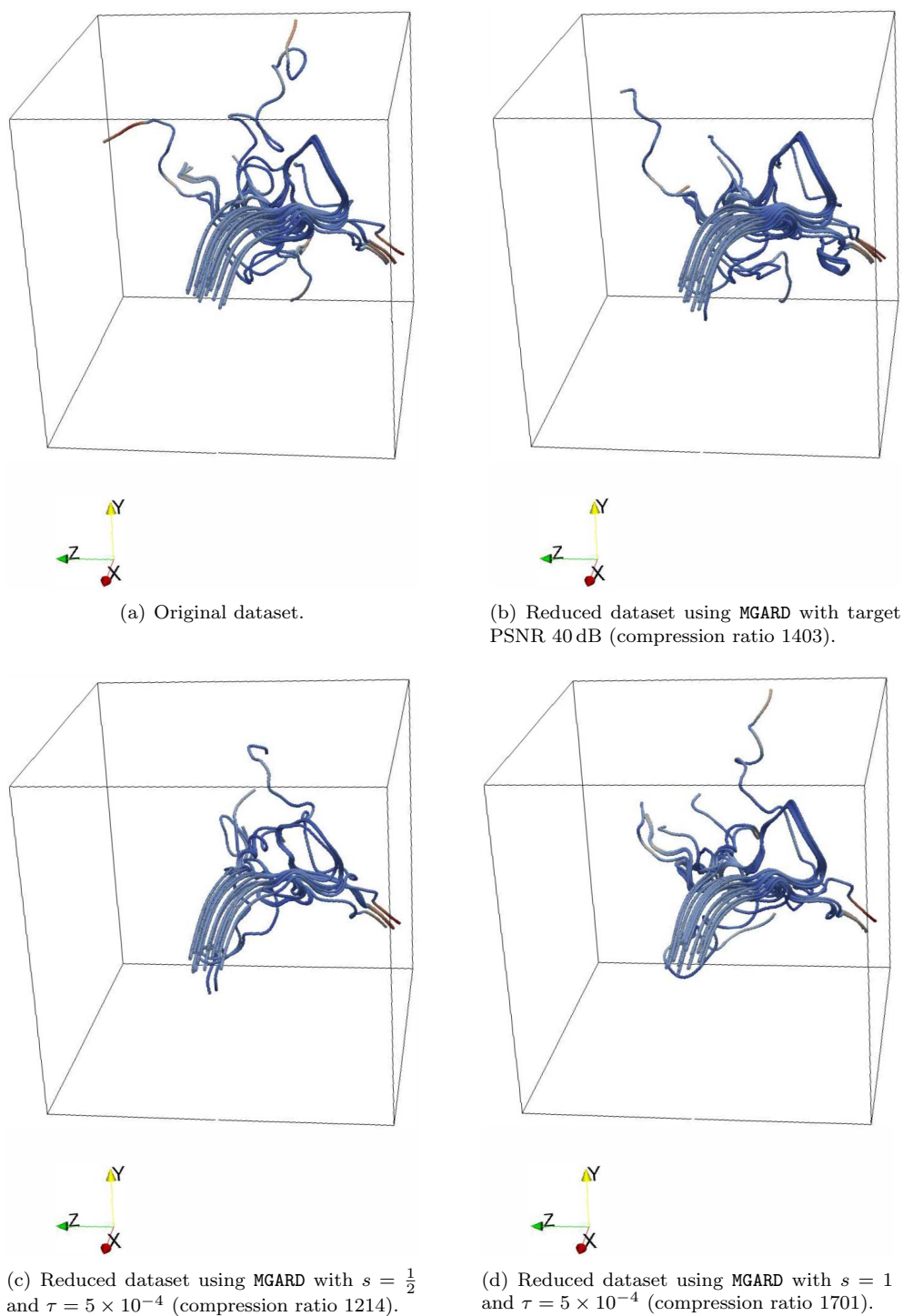
(a) Original dataset.

(b) Reduced dataset using `MGARD` with target PSNR 40 dB (compression ratio 1403).

(c) Reduced dataset using `MGARD` with $s = \frac{1}{2}$ and $\tau = 5 \times 10^{-4}$ (compression ratio 1214).

(d) Reduced dataset using `MGARD` with $s = 1$ and $\tau = 5 \times 10^{-4}$ (compression ratio 1701).

FIG. 9. *Streamlines generated from the velocity field of the turbulent combustion simulation for* 30 *initial points uniformly located on a disk of radius* 0.06 *centered at* $(1, 0.5, 0.5)$ *using the original dataset and reduced datasets produced by* `MGARD` *with different choices of smoothness parameter* s.
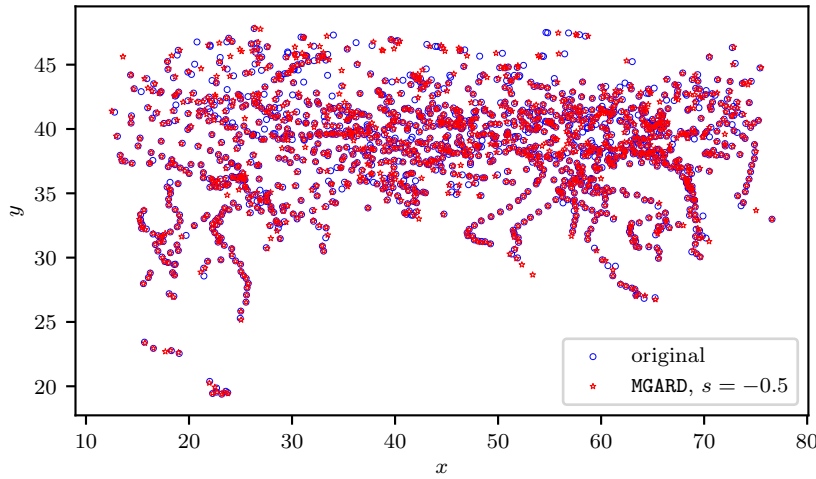
FIG. 10. *First* 1500 *blob centers as identified by* `openCV` *utilities. The circles are the blob centers found in the original dataset, and the stars are the blob centers found in a dataset reduced with* `MGARD` $(s = -\frac{1}{2})$ *with compression ratio* 8.5.

us to the smoothness parameter choice $s = -\frac{1}{2}$. The data is compressed by a factor of 8.5.

In Figure 10 we present the centers of the first 1500 blobs detected in the original and reduced datasets. It is evident from the figure that `MGARD` compression with $s = -\frac{1}{2}$ has minimal effect on the positioning of blob centers at the given compression ratio. Numerical experiments conducted using $s = 0$ and $s = -\frac{3}{2}$ have resulted in acceptable but slightly inferior (especially for $s = 0$) blob center positioning. To visualize, we plot frames 120–132 of the original and reduced ($s = -\frac{1}{2}$ and $s = 0$) datasets in four-frame increments in Figure 11. The blobs identified by `openCV` are marked.

6. Channel flow. As a final example, we consider data [11, 14, 19] obtained by a direct numerical simulation [13] of turbulent flow along a rectangular channel with $\text{Re}_\tau = 9.9935 \times 10^2$. The rectangular channel is of size $L_x \times L_y \times L_z = 8\pi \times 2 \times 3\pi$ (dimensionless units), and the discretization uses $N_x \times N_y \times N_z = 2048 \times 512 \times 1536$ nodes. The discretization is equispaced in the x and z directions, whereas a Chebyshev grid is employed in the wall-normal direction (y axis). The nonuniform spacing in the wall-normal direction can pose difficulties for other lossy compression approaches, such as wavelets. However, `MGARD` can handle any arbitrary spacing, including Chebyshev spacings, without any modifications. `MGARD` is used to compress the data from a single time instance, and each of the velocity components is compressed independently of the others. This dataset was also used in [2].

6.1. Spectrum control. In this subsection we will consider the effects of `MGARD` compression on the turbulent energy spectrum of the flow. Minimal or predictable distortion of the underlying energy spectrum is of paramount importance for turbulent flows, because this will allow the user to select or ascertain at what scale the flow structures are preserved. The energy spectrum is calculated by integrating the energy contained in the Fourier coefficients of the flow over spherical shells in Fourier space. Hence, the quantities of interest to be preserved are the Fourier coefficients of the
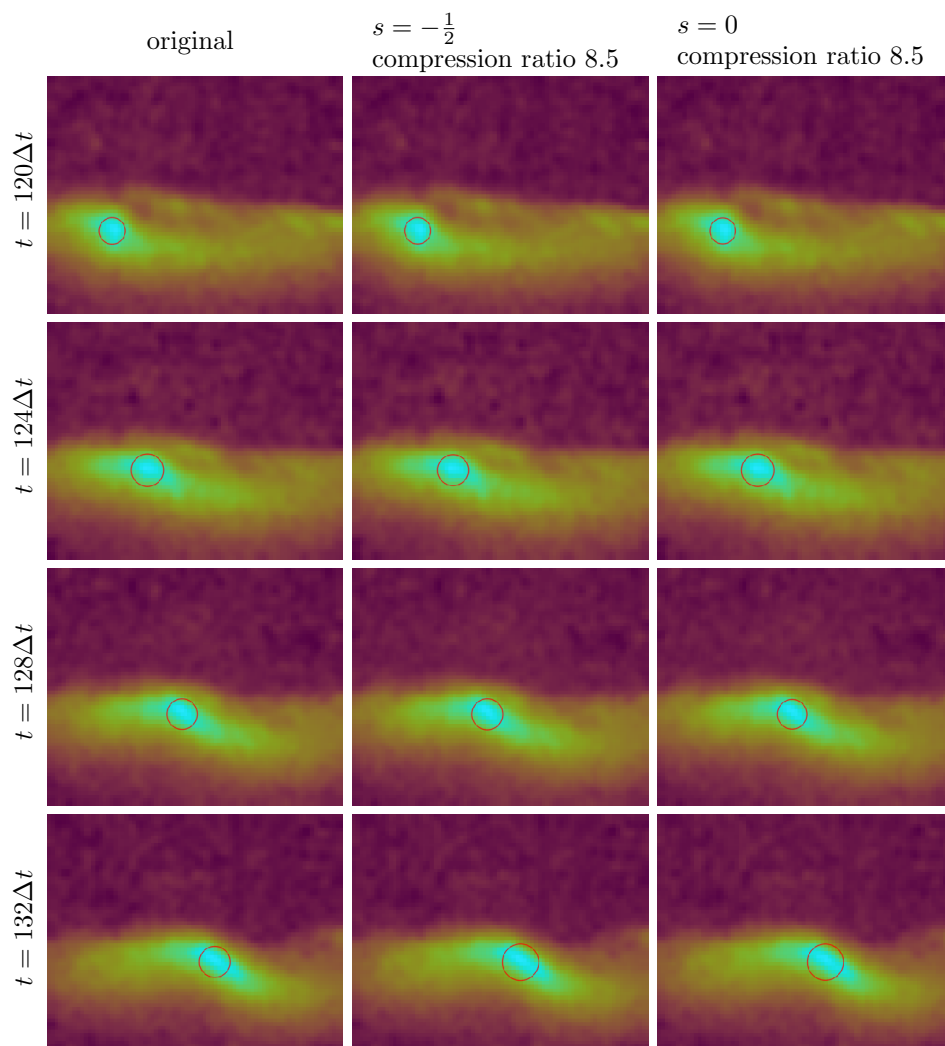
FIG. 11. *Blobs detected in the original and reduced datasets.*

velocity, given by

$$\mathcal{Q}_{\vec{k}}(u) = \int_\Omega u(\vec{x}) e^{i\vec{k}\cdot\vec{x}} \, d\vec{x}.$$

For τ a tolerance given, we would like to generate a reduced representation $\tilde{u}$ so that

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

for as many wavenumbers $\vec{k}$ as possible. To do this, we will apply `MGARD` with a smoothness parameter s and a tolerance ϵ on $\|u - \tilde{u}\|_s$ to be determined. In order to understand the effect of applying `MGARD` on the energy spectrum, we must determine which Fourier modes are resolved to an accuracy of τ when parameters s and ϵ are
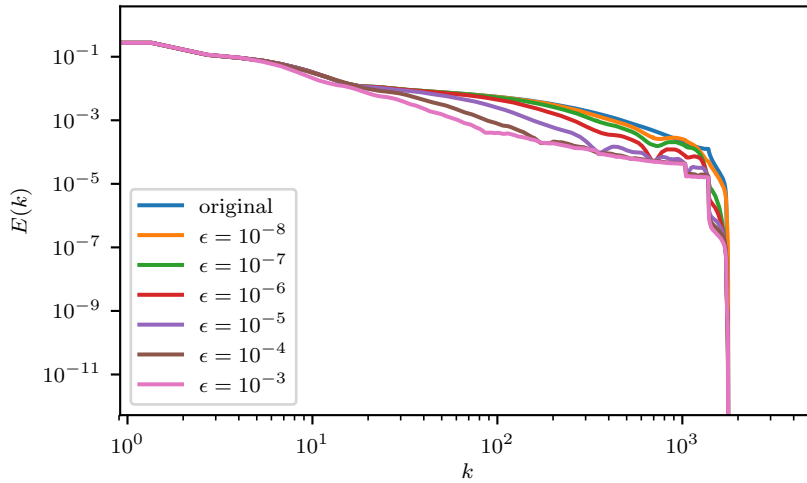
FIG. 12. *Effects of* `MGARD` *compression with* $s = -\frac{3}{2}$ *and various tolerances on the turbulent energy spectrum.*

used. It can be shown that (for $\vec{k} \neq \vec{0}$)

$$|\mathcal{Q}_{\vec{k}}(u) - \mathcal{Q}_{\vec{k}}(\tilde{u})| = |\mathcal{Q}_{\vec{k}}(u - \tilde{u})| = \left| \int_{\Omega} (u - \tilde{u})(x) \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}} \, \mathrm{d}\vec{x} \right|$$

$$\leq \|\vec{x} \mapsto \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}\|_{-s} \|u - \tilde{u}\|_{s} \lesssim |\vec{k}|^{-s} \|u - \tilde{u}\|_{s}.$$

Here $\vec{x} \mapsto \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}$ is the function returning the exponential $\mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}$ at the point $\vec{k}$, and $\|\vec{x} \mapsto \mathrm{e}^{\mathrm{i}\vec{k}\cdot\vec{x}}\|_{-s}$ is its norm as defined in (2) with index $-s$. If $\|u - \tilde{u}\|_{s} \lesssim \epsilon$, then any Fourier mode with $|\vec{k}| \leq (\tau/\epsilon)^{-1/s}$ will be satisfactorily resolved. As the exponential is smooth, we choose $s = -\frac{3}{2}$. Figure 12 illustrates `MGARD`'s impact on the turbulent energy spectrum of the channel flow field for various tolerances ϵ. As expected, as ϵ decreases the energy in the smaller scales is better preserved. Figure 13 displays the relationship between the tolerance ϵ set and the highest wavenumber $\vec{k}$ resolved.

7. Conclusion. This work detailed the theoretical framework enabling the data reduction package `MGARD` to perform adaptive compression of scientific data while preserving derived quantities. The primary goal of the project was to bridge the gap between the needs of application scientists, who interact with their data u mostly through derived quantities $\mathcal{Q}(u)$, and the capabilities of existing compression algorithms, which are generally geared toward preserving the data in $L^{\infty}(\Omega)$ (i.e., pointwise). In addition to the theory, the data structures and algorithms used in the implementation were presented. The method was applied to three datasets produced by numerical simulations and one dataset arising from a physical experiment. In each reduction, one of a family of quantities of interest, including mass averages, turbulent energy spectra, blob positioning, and streamlines, was preserved.

Appendix A. Lemmas for section 2. The results in this appendix concern the norm of $I - \Pi_{\ell-1}$ as an operator $L^2(V_\ell) \to L^2(V_\ell)$. We begin with a general lemma about norms of projections.
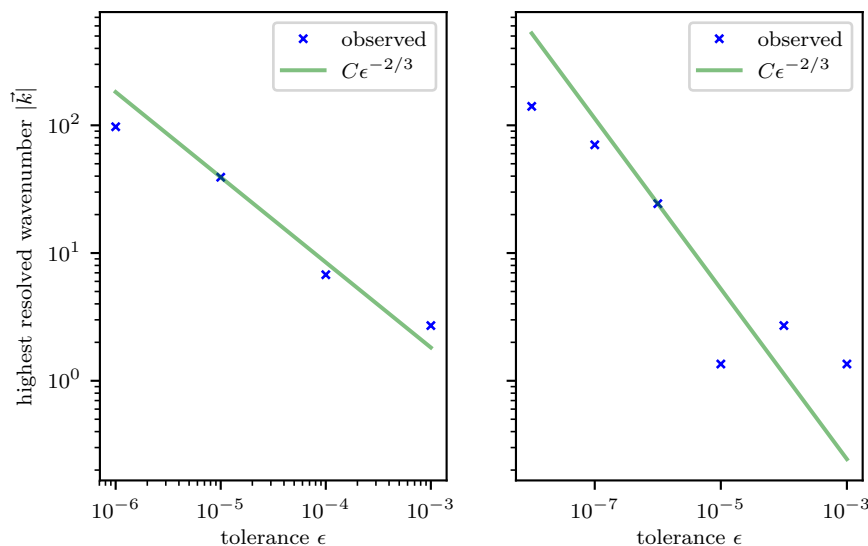
FIG. 13. *Highest resolved wavenumber* $|\vec{k}|$ *when using* `MGARD` *with* $s = -\frac{3}{2}$ *and various tolerances* ϵ *on* $\|u - \tilde{u}\|_s$. *The tolerance on the loss in the Fourier coefficients is* $\tau = 10^{-3}$ *in the left-hand subfigure and* $\tau = 10^{-4}$ *in the right-hand subfigure.*

LEMMA A.1. *Let X be a finite-dimensional inner product space. Let $X_0, X_1 \subset X$ be nonempty subspaces satisfying $X = X_0 \oplus X_1$. Let Π_0 and Π_1 denote the projections to X_0 and X_1, respectively. Then*

$$\|\Pi_0\| = \sup_{\substack{u \in X \\ \|u\|=1}} \|\Pi_0 u\|_X = \sup_{\substack{u \in X \\ \|u\|=1}} \|\Pi_1 u\|_X = \|\Pi_1\|,$$

where $\|\cdot\|_X$ is the norm given by the inner product $(\,\cdot\,,\,\cdot\,)_X$ and $\|\cdot\|$ is the operator norm.

Proof. Since $X_0 \cap X_1$ is trivial, there exists $\alpha \in [0,1)$ such that $|(u_0, u_1)_X| \leq \alpha \|u_0\|_X \|u_1\|_X$ for all $u_0 \in X_0$ and $u_1 \in X_1$. Let α be the best such constant. Then for all $u \in X$,

$$\begin{aligned}
\|u\|_X^2 &= \|\Pi_0 u\|_X^2 + 2(\Pi_0 u, \Pi_1 u)_X + \|\Pi_1 u\|_X^2 \\
&\geq \|\Pi_0 u\|_X^2 - 2\alpha \|\Pi_0 u\|_X \|\Pi_1 u\|_X + \|\Pi_1 u\|_X^2 \\
&\geq (1 - \alpha^2) \|\Pi_0 u\|_X^2,
\end{aligned}$$

so $\|\Pi_0 u\|_X \leq (1 - \alpha^2)^{-1/2} \|u\|_X$. As X is finite-dimensional, there exist unit vectors $e_0 \in X_0$ and $e_1 \in X_1$ satisfying $(e_0, e_1)_X = -\alpha$. Choose $u = e_0 + \alpha e_1$ so that $\Pi_0 u = e_0$ and $\Pi_1 u = \alpha e_1$. Then

$$\|u\|_X^2 = \|e_0\|_X^2 + 2\alpha(e_0, e_1)_X + \alpha^2 \|e_1\|_X^2 = (1 - \alpha^2) \|\Pi_0 u\|_X^2$$

and it follows that $\|\Pi_0\|_X^2 = (1 - \alpha^2)^{-1}$. The same argument applies if we interchange Π_0 and Π_1, so $\|\Pi_0\|_X = \|\Pi_1\|_X$. $\square$

In order to compute the operator norm of $I - \Pi_{\ell-1}$, we apply Lemma A.1 with $X = V_\ell$, $\Pi_0 = \Pi_{\ell-1}$, and $\Pi_1 = I - \Pi_{\ell-1}$. We first calculate $\Pi_{\ell-1}$ when $d = 1$.

LEMMA A.2. *Suppose $d = 1$, and let ρ be the mesh ratio of $\mathcal{P}_\ell$. Then $\|\Pi_{\ell-1}\| \leq \sqrt{\rho} + \sqrt{1/\rho}$, with equality holding if $\rho = 1$.*

Proof. Observe that

$$\|\Pi_{\ell-1}\|^2 = \sup_{\substack{v \in V_\ell \\ v \neq 0}} \frac{\|\Pi_{\ell-1}v\|^2}{\|v\|^2} = \sup_{\substack{v \in V_\ell \\ v \neq 0}} \frac{\sum_{K \in \mathcal{P}_{\ell-1}} \|\Pi_{\ell-1}v\|_K^2}{\sum_{K \in \mathcal{P}_{\ell-1}} \|v\|_K^2}.$$

Consider an individual element $K \in \mathcal{P}_{\ell-1}$. Let h^- and h^+ be the widths of the elements into which K is subdivided in $\mathcal{P}_\ell$. We may then, without loss of generality, assume that $K = (-h^-, h^+)$. Let $v \in V_\ell$ be arbitrary, and denote $v(-h^-)$, $v(0)$, and $v(h^+)$ by a, b, and c, respectively. Then

$$\begin{aligned}
\|\Pi_{\ell-1}v\|_K^2 &= \tfrac{1}{3}(h^- + h^+)(a^2 + ac + c^2) \\
&= \frac{1}{3}\frac{(h^- + h^+)^2}{h^- h^+}\big[h^- a^2 + h^- ab + (h^- + h^+)b^2 + h^+ bc + h^+ c^2\big] \\
&\quad - \frac{1}{3}\frac{h^- + h^+}{h^- h^+}\big[\tfrac{3}{4}(h^- a - h^+ c)^2 + \tfrac{1}{4}(h^- a + 2(h^- + h^+)b + h^+ c)^2\big] \\
(14) \qquad &\leq \frac{1}{3}\frac{(h^- + h^+)^2}{h^- h^+}\big[h^- a^2 + h^- ab + (h^- + h^+)b^2 + h^+ bc + h^+ c^2\big] \\
&= \frac{(h^- + h^+)^2}{h^- h^+}\|v\|_K^2 = (\sqrt{h^-/h^+} + \sqrt{h^+/h^-})^2\|v\|_K^2 \\
(15) \qquad &\leq (\sqrt{\rho} + \sqrt{1/\rho})^2\|v\|_K^2.
\end{aligned}$$

As a result,

$$(16) \qquad \|\Pi_{\ell-1}\|^2 \leq \sup_{\substack{v \in V_\ell \\ v \neq 0}} \frac{\sum_{K \in \mathcal{P}_{\ell-1}} (\sqrt{\rho} + \sqrt{1/\rho})^2\|v\|_K^2}{\sum_{K \in \mathcal{P}_{\ell-1}} \|v\|_K^2} = (\sqrt{\rho} + \sqrt{1/\rho})^2$$

and so $\|\Pi_{\ell-1}\| \leq \sqrt{\rho} + \sqrt{1/\rho}$.

Finally, consider the case of uniform grids. For each $K \in \mathcal{P}_{\ell-1}$, $h^- = h^+$. As a result, the inequality in (14) becomes an equality when $a = -2b = c$. Equality holds in (15) because $h^-/h^+ = h^+/h^- = \rho = 1$. The inequality in (16) becomes an equality because a single function $w \in V_\ell$ can maximize $\|\Pi_{\ell-1}w\|_K^2/\|w\|_K^2$ on all elements $K \in \mathcal{P}_{\ell-1}$ simultaneously. Indeed, let $w \in V_\ell$ be defined by $w(x) = 2$ for $x \in \mathcal{N}_{\ell-1}$ and $w(x) = -1$ for $x \in \mathcal{N}_\ell^*$. Then

$$\|\Pi_{\ell-1}\|^2 = \sup_{\substack{v \in V_\ell \\ v \neq 0}} \frac{\|\Pi_{\ell-1}v\|^2}{\|v\|^2} \geq \frac{\sum_{K \in \mathcal{P}_{\ell-1}} (\sqrt{1} + \sqrt{1})^2\|w\|_K^2}{\sum_{K \in \mathcal{P}_{\ell-1}} \|w\|_K^2} = 2^2. \qquad \square$$

Now we can extend to the case of general dimension to obtain the desired bound.

COROLLARY A.3. *Let $\mathcal{P}_\ell$ be the tensor product of d one-dimensional meshes $\{\mathcal{P}_\ell^i : 1 \leq i \leq d\}$ with mesh ratios $\{\rho_i : 1 \leq i \leq d\}$. Then $\|I - \Pi_{\ell-1}\| \leq \prod_{i=1}^d \sqrt{\rho_i} + \sqrt{1/\rho_i}$, with equality holding if $\rho_i = 1$ in each dimension.*

Proof. Because $\mathcal{P}_\ell$ is a tensor product mesh, the interpolation from $\mathcal{P}_\ell$ to $\mathcal{P}_{\ell-1}$ is given by the tensor product of the one-dimensional interpolations from $\mathcal{P}_\ell^i$ to $\mathcal{P}_{\ell-1}^i$: $\Pi_{\ell-1} = \Pi_{\ell-1}^1 \otimes \cdots \otimes \Pi_{\ell-1}^d$. By the definition of the tensor product, then, $\|\Pi_{\ell-1}\| = \|\Pi_{\ell-1}^1\| \cdots \|\Pi_{\ell-1}^d\|$, and the result follows by Lemma A.2. $\qquad\square$

Appendix B. Regularity calculation for subsection 4.3. The streamline discussion in subsection 4.3 involves integration of a difference $u - \tilde{u}$ of flow fields along a path Γ. Because in practice that integration is carried out numerically, it suffices to consider piecewise linear paths.

LEMMA B.1. *Let* $\Gamma : [0, t] \to \Omega \subset \mathbb{R}^d$ *be a piecewise linear path, and let* χ_Γ *be the indicator function of the trajectory of* Γ. *If* $d = 1$, $\chi_\Gamma \in H^r(\Omega)$ *for* $r < 1/2$. *Otherwise,* $\chi_\Gamma \in H^r(\Omega)$ *for* $r < (1 - d)/2$.

Proof. It suffices to consider the case that Γ consists of a single linear segment with parametrization $\gamma : [0, 1] \to \Omega$. Without loss of generality, assume that $\gamma(0) = \vec{0}$ and $\gamma(1) = \vec{e}_1$. Let χ_γ be the characteristic function of the trajectory of γ. The Fourier coefficients of χ_γ are given by

$$(17) \qquad \hat{\chi}_\gamma(\omega) = \int_\Omega e^{i(\omega \cdot x)} \chi_\gamma(x) \, dx = \int_0^1 1 e^{i(\omega \cdot \zeta \vec{e}_1)} \, d\zeta = \frac{1}{i(\omega \cdot \vec{e}_1)} \left[e^{i(\omega \cdot \vec{e}_1)} - 1 \right]$$

if $\omega \cdot \vec{e}_1$ is nonzero. Otherwise, the coefficients are given by

$$\hat{\chi}_\gamma(\omega) = \int_\Omega e^{i(\omega \cdot x)} \chi_\gamma(x) \, dx = \int_\Omega 1 \chi_\gamma(x) \, dx = 1.$$

As 1 is the limiting value of (17) as $\omega \cdot \vec{e}_1 \to 0$, we will use (17) as our expression for $\hat{\chi}_\gamma(\omega)$, with the understanding that its value is 1 when $\omega \cdot \vec{e}_1 = 0$. In order for χ_γ to be included in the Sobolev space $H^r(\Omega)$, its Fourier coefficients must satisfy the summability condition

$$(18) \qquad \sum_{\omega \in \mathbb{Z}^d} |\hat{\chi}_\gamma(\omega)|^2 (1 + |\omega|^2)^r < \infty.$$

The square moduli of the Fourier coefficients are given by

$$|\hat{\chi}_\gamma(\omega)|^2 = \frac{1}{-i^2 (\omega \cdot \vec{e}_1)^2} \left[e^{i(\omega \cdot \vec{e}_1)} - 1 \right] \left[e^{-i(\omega \cdot \vec{e}_1)} - 1 \right] = 2 \frac{1 - \cos(\omega \cdot \vec{e}_1)}{(\omega \cdot \vec{e}_1)^2}$$

with the understanding that the value of this expression is 1 when $\omega \cdot \vec{e}_1 = 0$. Let ω_1 denote the first coordinate of ω, so that $\omega \cdot \vec{e}_1 = \omega_1$. Observe that

$$(19) \qquad 2 \frac{1 - \cos(\omega_1)}{\omega_1^2} \leq \begin{cases} 4/\omega_1^2, & \omega_1 \neq 0, \\ 1, & \omega_1 = 0, \end{cases}$$

with equality holding in the $\omega_1 = 0$ case. Following this bound, we can split the sum in (18) as follows:

$$(20) \qquad \underbrace{\sum_{\substack{\omega \in \mathbb{Z}^d \\ \omega_1 = 0}} |\hat{\chi}(\omega)|^2 (1 + |\omega|^2)^r}_{(20.\text{i})} + \underbrace{\sum_{\substack{\omega \in \mathbb{Z}^d \\ \omega_1 \neq 0}} |\hat{\chi}(\omega)|^2 (1 + |\omega|^2)^r}_{(20.\text{ii})}.$$

Consider first (20.i), the left-hand sum. Let ω' denote the vector $(\omega_2, \ldots, \omega_d)$. It can be shown that the number of vectors $\omega' \in \mathbb{Z}^{d-1}$ with modulus R is at most $O(R^{d-2})$. Then, by (19),

$$(20.\text{i}) = \sum_{\substack{\omega \in \mathbb{Z}^d \\ \omega_1 = 0}} |\hat{\chi}(\omega)|^2 (1 + |\omega|^2)^r = \sum_{\omega' \in \mathbb{Z}^{d-1}} 1(1 + |\omega'|^2)^r \lesssim \sum_{R=0}^{\infty} (1 + R^2)^r R^{d-2},$$

which is finite if $d - 2 + 2r < -1$, or $r < (1-d)/2$. Consider now (20.ii), the right-hand sum. Using (19),

$$(20.\text{ii}) = \sum_{\substack{\omega \in \mathbb{Z}^d \\ \omega_1 \neq 0}} |\hat{\chi}(\omega)|^2 (1 + |\omega|^2)^r \leq \sum_{\omega' \in \mathbb{Z}^{d-1}} \sum_{\substack{\omega_1 \in \mathbb{Z} \\ \omega_1 \neq 0}} \frac{4}{\omega_1^2} (1 + |\omega_1|^2 + |\omega'|^2)^r.$$

Consider the inner sum. We may bound it by an integral as follows:

$$\sum_{\substack{\omega \in \mathbb{Z} \\ \omega \neq 0}} \frac{4}{\omega_1^2} (1 + |\omega_1|^2 + |\omega'|^2)^r \leq 8(2 + |\omega'|^2)^r + 8 \int_1^{\infty} \frac{1}{\omega_1^2} (1 + |\omega_1|^2 + |\omega'|^2)^r \, \mathrm{d}\omega_1.$$

Define a substitution variable μ by $\mu = \omega_1 / \sqrt{1 + |\omega'|^2}$. Then $\omega_1^2 = \mu^2(1 + |\omega'|^2)$ and $1 + |\omega|^2 = (1 + \mu^2)(1 + |\omega'|^2)$, so that, with $L(\omega')$ denoting $1/\sqrt{1 + |\omega'|^2}$,

$$(21) \qquad \int_1^{\infty} \frac{1}{\omega_1^2} (1 + |\omega_1|^2 + |\omega'|^2)^r \, \mathrm{d}\omega_1 = \int_{L(\omega')}^{\infty} \frac{(1 + \mu^2)^r}{\mu^2} (1 + |\omega'|^2)^{r-1/2} \, \mathrm{d}\mu.$$

We pull out the constant factor $(1 + |\omega'|^2)^{r-1/2}$ and restrict our attention to the case $r < 0$. Bounding $1 + \mu^2$ below by 1 on $[L(\omega'), 1]$ and by μ^2 on $[1, \infty)$, we can estimate the integral as follows:

$$\int_{L(\omega')}^{\infty} \frac{(1 + \mu^2)^r}{\mu^2} \, \mathrm{d}\mu = \int_{L(\omega')}^{1} \frac{(1 + \mu^2)^r}{\mu^2} \, \mathrm{d}\mu + \int_1^{\infty} \frac{(1 + \mu^2)^r}{\mu^2} \, \mathrm{d}\mu$$

$$\leq \int_{L(\omega')}^{1} \mu^{-2} \, \mathrm{d}\mu + \int_1^{\infty} \mu^{2r-2} \, \mathrm{d}\mu$$

$$= \sqrt{1 + |\omega'|^2} - 1 + \frac{1}{2r - 1} \leq \sqrt{1 + |\omega'|^2}.$$

Substituting into the expression for (20.ii), we find that

$$(20.\text{ii}) \leq \sum_{\omega' \in \mathbb{Z}^{d-1}} 8(2 + |\omega'|^2)^r + 8(1 + |\omega'|^2)^r,$$

which is finite so long as $d - 2 + 2r < -1$, or $r < (1 - d)/2$. We conclude that (18) holds, and therefore $\chi_\Gamma \in H^r(\Omega)$, if $r < (1 - d)/2$.

Special attention is warranted for the case $d = 1$, as the above analysis is pessimistic then. Observe that (20.i) is finite independent of r in this case. The remaining

sum, (20.ii), is given by

$$(20.\text{ii}) = \sum_{\substack{\omega_1 \in \mathbb{Z} \\ \omega_1 \neq 0}} |\hat{\chi}(\omega)|^2 \, (1 + |\omega_1|^2)^s = \sum_{\substack{\omega_1 \in \mathbb{Z} \\ \omega_1 \neq 0}} \frac{4}{\omega_1^2} \, (1 + |\omega_1|^2)^r,$$

which is finite as long as $2r - 2 < -1$, or $r < \frac{1}{2}$. So, in one dimension $\chi_\Gamma \in H^r(\Omega)$ for all $r < \frac{1}{2}$.   □

REFERENCES

[1] M. Ainsworth, S. Klasky, and B. Whitney, *Compression using lossless decimation: Analysis and application*, SIAM J. Sci. Comput., 39 (2017), pp. B732–B757.

[2] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, *Multilevel techniques for compression and reduction of scientific data—the multivariate case*, SIAM J. Sci. Comput., (2018).

[3] M. Ainsworth, O. Tugluk, B. Whitney, and S. Klasky, *Multilevel techniques for compression and reduction of scientific data—the univariate case*, Comput. Vis. Sci., 19 (2018), pp. 65–76.

[4] L. A. Bautista Gomez and F. Cappello, *Improving floating point compression through binary masks*, in Proceedings of the 2013 IEEE International Conference on Big Data, 2013, pp. 326–331.

[5] F. Bornemann and H. Yserentant, *A basic norm equivalence for the theory of multilevel methods*, Numer. Math., 64 (1993), pp. 455–476.

[6] G. Bradski, *The OpenCV library*, Dr. Dobb's Journal of Software Tools, 2000.

[7] S. Di and F. Cappello, *Fast error-bounded lossy HPC data compression with SZ*, in Proceedings of the 2016 IEEE 30th International Parallel and Distributed Processing Symposium, Chicago, IL, 2016, pp. 730–739.

[8] J. Diffenderfer, A. Fox, J. Hittinger, G. Sanders, and P. Lindstrom, *Error Analysis of ZFP Compression for Floating-Point Data*, arXiv:1805.00546, 2018.

[9] D. A. D'Ippolito, J. R. Myra, and S. J. Zweben, *Convective transport by intermittent blob-filaments: Comparison of theory and experiment*, Phys. Plasmas, 18 (2011), 060501.

[10] D. L. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.

[11] *Johns Hopkins Turbulence Databases*, Channel Flow Dataset Description, 2017, http://turbulence.pha.jhu.edu.

[12] H. Kolla, X.-Y. Zhao, J. H. Chen, and N. Swaminathan, *Velocity and reactive scalar dissipation spectra in turbulent premixed flames*, Combust. Sci. Technol., 188 (2016), pp. 1424–1439.

[13] M. Lee and R. D. Moser, *Direct numerical simulation of turbulent channel flow up to $Re_\tau \approx$ 5200*, J. Fluid Mech., 774 (2015), pp. 395–415.

[14] Y. Li, E. Perlman, M. Wan, Y. Yang, C. Meneveau, R. Burns, S. Chen, A. Szalay, and G. Eyink, *A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence*, J. Turbul., 9 (2008), N31.

[15] P. Lindstrom and M. Isenburg, *Fast and efficient compression of floating-point data*, IEEE Trans. Vis. Comput. Graph., 12 (2006), pp. 1245–1250.

[16] R. J. Maqueda, G. A. Wurden, D. P. Stotler, S. J. Zweben, B. LaBombard, J. L. Terry, J. L. Lowrance, V. J. Mastrocola, G. F. Renda, D. A. DÍppolito, J. R. Myra, and N. Nishino, *Gas puff imaging of edge turbulence (invited)*, Rev. Sci. Instrum., 74 (2003), pp. 2020–2026.

[17] J. R. Myra, W. M. Davis, D. A. D'Ippolito, B. LaBombard, D. A. Russell, J. L. Terry, and S. J. Zweben, *Edge sheared flows and the dynamics of blob-filaments*, Nuclear Fusion, 53 (2013), 073013.

[18] P. Oswald, *Multilevel Finite Element Approximation: Theory and Applications*, Teubner Skripten Numer., Teubner, Stuttgart, 1994.

[19] E. Perlman, R. Burns, Y. Li, and C. Meneveau, *Data exploration of turbulence simulations using a database cluster*, in Proceedings of the 2007 ACM/IEEE Conference on Supercomputing, Vol. 23, Reno, NV, 2007.

[20] E. M. Stein and R. Shakarchi, *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*, Princeton University Press, Princeton, NJ, 2005.

[21] D. Tao, S. Di, Z. Chen, and F. Cappello, *Significantly improving lossy compression for scientific data sets based on multidimensional prediction and error-controlled quantization*, in Proceedings of the IEEE International Parallel and Distributed Processing Symposium, Orlando, FL, 2017, pp. 1129–1139.

[22] E. H. Twizell, A. B. Gumel, and Q. Cao, *A second-order scheme for the "Brusselator" reaction–diffusion system*, J. Math. Chem., 26 (1999), pp. 297–316.

[23] Z. Wang and A. C. Bovik, *Mean squared error: Love it or leave it?*, IEEE Signal Process., 26 (2009), pp. 98–117.

[24] L. Wu, K. J. Wu, A. Sim, M. Churchill, J. Y. Choi, A. Stathopoulos, C. S. Chang, and S. Klasky, *Towards real-time detection and tracking of spatio-temporal features: Blob-filaments in fusion plasma*, IEEE Trans. Big Data, 2 (2016), pp. 262–275.

[25] S. J. Zweben, J. R. Myra, W. M. Davis, D. A. Díppolito, T. K. Gray, S. M. Kaye, B. P. LeBlanc, R. J. Maqueda, D. A. Russell, and D. P. Stotler, *Blob structure and motion in the edge and SOL of NSTX*, Plasma Phys. Control. Fusion, 58 (2016), 044007.