Velocity gradient prediction using parameterized Lagrangian deformation models

Criston Hyett, Yifeng Tian, Michael Woodward, Misha Stepanov, Chris Fryer, Michael Chertkov, Daniel Livescu

July 30, 2023

Abstract

We seek to efficiently predict the statistical evolution of the velocity gradient tensor (VGT) by creating local models for the pressure Hessian. Previous work has identified physics-informed machine learning (PIML) to be adept in this prediction; of note in this class of models is the Tensor Basis Neural Network (TBNN) for its embedded physical constraints and demonstrated performance. Simultaneously, phenomenological models were advanced by approximating the local closure to the pressure Hessian via deformation models using the history of the VGT. The latest in this series of models is the Recent Deformation of Gaussian Fields (RDGF) model. In this work, we combine the (local in time) PIML approach with the phenomenological idea of inclusion of recent deformation to create a data-driven Lagrangian deformation model. We compare the model performance to both the TBNN and the RDGF models, and provide data-driven hypotheses regarding the upstream assumptions made in the RDGF model.

1 Introduction

The velocity gradient tensor (VGT) describes many important aspects of turbulence. It displays characteristic non-Gaussian statistics including intermittency, describes the deformation rate of a fluid volume, and contains encapsulates alignment between strain and vorticity [27]. The VGT has been the subject of much study, and with the exponential growth of compuational power, direct numerical simulations that resolve the smallest scales of turbulence have provided enormous VGT datasets. Even more recently, machine learning (ML), particularly physics-informed machine learning (PIML), have shown remarkable ability to glean predictive capability from these large datasets, suggesting patterns exist that we have not recognized yet.

Of particular note in recent years is recent deformation of Gaussian fields (RDGF) [22], wherein the authors postulate a closure of the VGT evolution equations via choosing a Gaussian upstream condition that is deformed according to the current VGT. This hypothesis is along the same lines as the authors' attempt to close the equations using deformation of a small fluid element [6], but taken to the limit of small time and space.

In the realm of machine learning, the work is inspired by Tian et.al[41], that used a highly structured, so-called tensor basis neural network architecture to close the equations. This network structure was inspired by theoretical work by [25], [22], [31].

Our goals in this paper are to advance the phenomenology via data analysis, and leverage this to improve the TBNN methodology to better predict the statistical evolution of the VGT.

2 Previous Work

2.1 Governing Equations for VGT

Navier-Stokes defines the evolution of a velocity field given a pressure field

$$\frac{\partial u_i}{\partial t} + u_k \frac{\partial u_i}{\partial x_k} = -\frac{\partial P}{\partial x_i} + \nu \frac{\partial^2 u_i}{\partial x_k \partial x_k} \tag{1}$$

The VGT is defined by $A_{ij} = \frac{\partial u_i}{\partial x_j}$, so we apply spatial derivatives to eq(1), and use the definition of material derivative to obtain an ODE for the velocity gradient tensor (defined in the Lagrangian frame)

$$\frac{dA_{ij}}{dt} = \frac{\partial A_{ij}}{\partial t} + u_k \frac{\partial A_{ij}}{\partial x_k} = -A_{ik}A_{kj} - \frac{\partial^2 P}{\partial x_i \partial x_j} + \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k} \tag{2}$$

Using the incompressibility condition,

$$\nabla \cdot u = 0 \implies A_{ii} = 0 \tag{3}$$

we can take the trace of eq(2) to find

$$\frac{\partial^2 P}{\partial x_k \partial x_k} = -A_{ij}A_{ji} \tag{4}$$

which determines the trace of the pressure Hessian. Thus, we can define the so-called deviatoric pressure Hessian as

$$H_{ij} := -\left(\frac{\partial^2 P}{\partial x_i \partial x_j} - \frac{1}{3} \frac{\partial P}{\partial x_k \partial x_k} \delta_{ij} \right) \tag{5}$$

Viellefosse [42] and Cantwell [3] studied the purely local "Restricted Euler" dynamics:

$$E_{ij} := -\left(A_{ik}A_{kj} - \frac{1}{3} A_{mn}A_{nm}\delta_{ij} \right) \tag{6}$$

Finally, letting the viscous term be denoted by

$$T_{ij} := \nu \frac{\partial^2 A_{ij}}{\partial x_k \partial x_k} \tag{7}$$

We can write the ODE for the Lagrangian VGT as

$$\frac{dA_{ij}}{dt} = E_{ij} + H_{ij} + T_{ij} \tag{8}$$

Via numerical studies, the main challenge is to predict the deviatoric pressure Hessian.

2.2 Tensor Basis Neural Network

We can write the formal (nonlocal) solution for the deviatoric pressure hessian as[29]

$$H_{ij}(\mathbf{x}) = \iiint \frac{\delta_{ij} - \hat{r_i}\hat{r_j}}{2\pi r^3} Q(\mathbf{x}+\mathbf{r}) d\mathbf{r} \tag{9}$$

The challenge then, is to obtain an approximation of this nonlocal integral using only local information.

Following Lawson & Dawson[25], an expansion of the integral is proposed, first as a Taylor series expansion in $Q(x+r)$

$$\hat{H} = \sum_{m,n=0}^{\infty} \alpha_{mn} S^m W^n \tag{10}$$

where

$$S = \frac{1}{2}(A + A^T) \qquad W = \frac{1}{2}(A - A^T) \tag{11}$$

Finally we can reduce from an infinite sum using Cayley-Hamilton, and expand via the tensor basis[44],[31]

$$\hat{H} = \sum_{n=1}^{10} g^{(n)}(\lambda_1, \ldots, \lambda_5) T^{(n)} \tag{12}$$

with $g^{(n)}$ scalar functions of the invariants

$$\lambda_1 = \mathrm{tr}(S^2) \quad \lambda_2 = \mathrm{tr}(W^2) \quad \lambda_3 = \mathrm{tr}(S^3) \quad \lambda_4 = \mathrm{tr}(W^2 S) \quad \lambda_5 = \mathrm{tr}(W^2 S^2) \tag{13}$$

and the tensor basis given by:

$$T^{(1)} = S \qquad T^{(2)} = SW - WS \tag{14}$$

$$T^{(3)} = S^2 - \frac{1}{3} I \cdot \mathrm{tr}(S^2) \qquad T^{(4)} = W^2 - \frac{1}{3} I \cdot \mathrm{tr}(W^2) \tag{15}$$

$$T^{(5)} = WS^2 - S^2 W \qquad T^{(6)} = W^2 S + SW^2 - \frac{2}{3} I \cdot \mathrm{tr}(SW^2) \tag{16}$$

$$T^{(7)} = WSW^2 - W^2 SW \qquad T^{(8)} = SWS^2 - S^2 WS \tag{17}$$

$$T^{(9)} = W^2 S^2 + S^2 W^2 - \frac{2}{3} I \cdot \mathrm{tr}(S^2 W^2) \quad T^{(10)} = WS^2 W^2 - W^2 S^2 W \tag{18}$$

This formulation reduces the challenge to finding the functions of known scalars, i.e., learning the $g^{(n)}$'s as shown in fig(1).

This network architecture ensures the output tensor $\hat{H}$ is symmetric, traceless, and the network itself is rotationally and Galilean invariant.

Tian et.al[41] showed ability to train the network, and demonstrated state-of-the-art performance on a variety of relevant physical metrics: eigenvector alignments, Q-R conditional mean tangents (CMTs), *a posteriori* tests evolving an initially Gaussian distributed VGT field to fully developed turbulence as evidenced by characteristic teardrop-shaped Q-R probability distributions (PDFs).

In particular, Tian et.al used

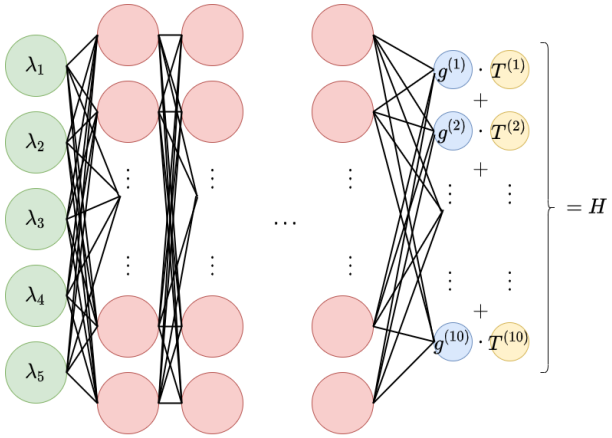

Figure 1: The architecture of the TBNN. The invariants and tensor basis elements are calculated from each sample of the VGT, the invariants are then used as input to a fully connected network, and the resulting $g^{(i)}$ multiply the corresponding tensor basis elements $T^{(i)}$, before summing into the prediction $\hat{H}$

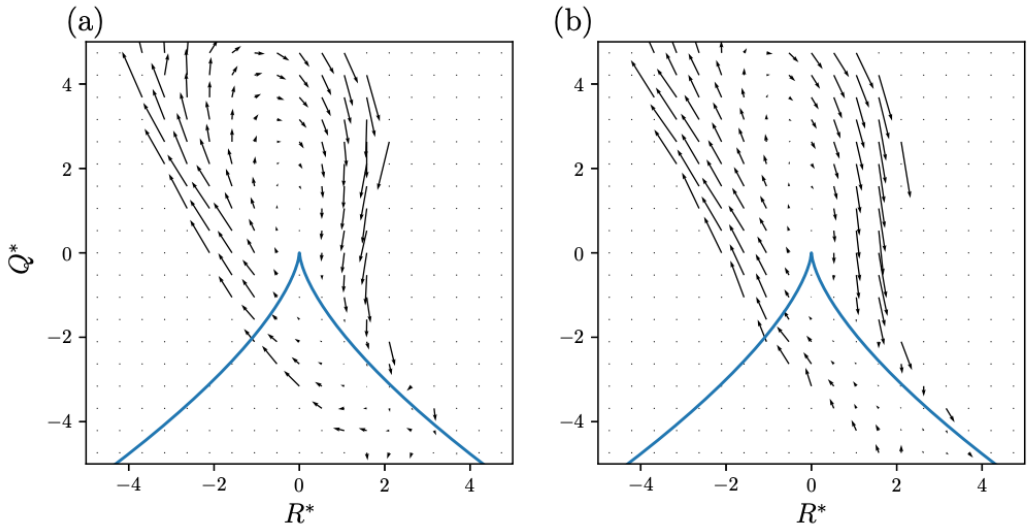

Figure 2: Results from Tian et.al[41] showing Q-R CMTs from (a) DNS data and (b) trained TBNN

2.3 Recent Deformation of Gaussian Fields

A note on notation in this section - I use the notation of Johnson & Meneveau as I am explaining their model, and they mean something specific by using it. Throughout the rest of the paper, the idea is however much the same, e.g., a neural network attempts to make the best approximation of the expected pressure Hessian, conditioned upon samples. Thus the change in notation is largely artificial here.

Johnson and Meneveau[22] introduced the RDGF model following a long line of postulating that the deviatoric part of the pressure Hessian could be locally modeled using information of the time history of the deformation tensor [6],[8],[7].

Chertkov et.al's Tetrad Model, attempted to capture this deformation directly by following a small fluid element approximated by 4 Lagrangian particles. This model avoided the finite time singularity resulting from using only the restricted Euler term[3], and attempts to fuse this low-dimensional model of a fluid element with data-driven methods for prediction of the coarse-grained VGT are ongoing[17],[16]. Chevillard & Meneveau used the same initially isotropic upstream condition as in the Tetrad Model when developing their Recent Fluid Deformation (RFD) model. Finally, Johnson & Meneveau enriched the approximation of the upstream condition using high fidelity direct numerical simulation (DNS) data[43].

RDGF models the conditional average of the pressure Hessian as

$$\langle P_{ij}|A\rangle = \left\langle \frac{\partial^2 p}{\partial x_i \partial x_j}|A \right\rangle = \frac{\partial X_k}{\partial x_i}\left\langle \frac{\partial^2 p}{\partial X_k \partial X_l}|A \right\rangle \frac{\partial X_l}{\partial x_j} = D^{-1}_{ki}\langle \tilde{P}_{kl}|A\rangle D^{-1}_{lj} \tag{19}$$

Where $\tilde{P_{kl}}$ is an upstream pressure Hessian, and the deformation tensor is given by

$$D_{ij} = \frac{\partial x_i}{\partial X_j} \tag{20}$$

and evolves according to (assuming isotropic initial condition - this is without loss of generality for sufficiently far removed upstream times)

$$\frac{dD_{ij}}{dt} = A_{ik}D_{kj} \quad \text{with } D_{ij}(0) = \delta_{ij} \tag{21}$$

so that the general solution is defined via the time-ordered exponential

$$D_{ij}(t) = \mathrm{Texp}(A(t)) = \lim_{N\to\infty}\prod_{i=0}^{N}\left(e^{A(t_i)\Delta t}\right) \tag{22}$$

where the product is the left product, and $t_0 = 0, t_N = t$. If we take the leading order term, we get the RDGF approximation

$$D(x,t) \approx \exp\left[A(x,t)\Delta t\right] \tag{23}$$

The upstream pressure Hessian is given by

$$\langle \tilde{P}_{ij}|A\rangle \approx \frac{1}{3}\langle \tilde{P}_{kk}|A\rangle\delta_{ij} + \langle \tilde{P}^{(d)}_{ij}|A\rangle_{\text{Gaussian}} \tag{24}$$

where, using the notation of the tensor basis, eqs(14-18)

$$\langle \tilde{P}^{(d)}_{ij}|A\rangle_{\text{Gaussian}} = \gamma T^{(2)} + \alpha T^{(3)} + \beta T^{(4)} \tag{25}$$

with the coefficients

$$\alpha = -\frac{2}{7}, \quad \beta = -\frac{2}{5}, \quad \gamma \approx 0.08 \tag{26}$$

By enforcing the prediction eq(24) to have trace of $2Q$, and labeling

$$G_{ij} := D_{mi}^{-1} \langle \tilde{P}_{ij}^{(d)} | A \rangle_{\text{Gaussian}} \tag{27}$$

The prediction of the pressure Hessian becomes

$$\langle P_{ij} | A \rangle = 2Q \frac{C_{ij}^{-1}}{C_{kk}^{-1}} + G_{ij} - \frac{C_{ij}^{-1}}{C_{kk}^{-1}} G_{ll} \tag{28}$$

3 Methodology

This section is a work-in-progress. I talk more about this in the prospectus, and welcome comments and ideas.

3.1 Proposed Models

Our approach is to utilize the idea of biasing the statistics of the conditional deviatoric pressure Hessian using the history of the VGT (as in the RDGF, RFD, and Tetrad models), while avoiding postulating the exact functional form; instead allowing it to be data-driven using PIML (as in TBNN model).

We do this by augmenting the inputs to the feed-forward portion of the TBNN, the idea being to bias the statistics of the invariants, while believing that the tensors derived from the local-in-time VGT are sufficient to span the function space.

3.1.1 Temporal Convolution

The simplest of our proposed models augments the set of invariants using a learned temporal convolution.

$$\hat{H} = \sum_{n=1}^{10} g_\theta^{(n)}(\lambda_1, \ldots, \lambda_5, X) T^{(n)} \tag{29}$$

with

$$X(t) = \sum_{t_f}^{t_f - \Delta t \cdot n} \zeta_n A_n \tag{30}$$

where ζ_n are learned simultaneously with the parameterization of the scalar g functions.

This model avoids biasing the prediction with phenomenological thoery, but at the cost of computational overhead. However, because of the simple structure, the learned ζ's may inform a temporal correlation in the VGT that is useful in the prediction of the deviatoric pressure Hessian. An extreme case of very short time correlations would reinforce the hypothesis presented in RDGF, while longer temporal correlations would support the modeling approaches of e.g., [40].

3.1.2 Time-Ordered Exponential

Finally, we may take an intermediate step between the two, and introduce the time-ordered exponential to the network, setting

$$X(t) = \text{Texp}[A] \tag{31}$$

in eq(29)

3.2 Data Analysis

In this section we perform numerical studies on DNS data to evaluate assumptions and choices for model hyperparameters.

3.2.1 Ground Truth Data

Our data is DNS of forced isotropic turbulence with $Re = 240$. Eularian data is generated, random points are sampled, VGT and pressure Hessian are constructed at these points using interpolation, and Lagrangian trajectories of these quantities are advanced using the Eularian velocity fields. We have 122k samples, each advanced for 1000 timesteps with $\Delta t = 3e - 4$, representing approximately half an inertial eddy turnover time. For each sample and each timestep, we have measured VGT, PH, and viscous terms.

3.2.2 Cramer Functions of Strain Rate Eigenvalues

Coarsely stated, the Cramer function (also more generally called the rate function) determines how quickly large deviations from a mean diminish. If we believe in the ability to close the equations for the deviatoric pressure Hessian locally, it is apparent that there must exist a relation between temporal correlations of the strain-rate tensor, and temporal correlations of the deviatoric pressure Hessian.

3.2.3 Gaussian Upstream Condition

The Gaussian upstream condition present in RDGF is convenient, but for short deformation time is known to be inaccurate. Via analysis of Lagrangian data, we can evaluate how long the deformation time must be to transform the Gaussian upstream condition to a distribution representative of fully developed turbulence. In particular, we can evaluate

$$\min_{\Delta t} \left\| P_{gt}^{(d)} - P'^{(d)}_{\text{Gaussian}}(\Delta t) \right\| \tag{32}$$

where $P'^{(d)}_{\text{Gaussian}}(\Delta t)$ is the true deformation of the Gaussian upstream condition, by the ground truth VGT over a time Δt.

4 Results

4.1 Hyperparameter Tuning

One important deviation from the guiding paper is in data normalization. Tian et.al, normalized the VGT and PH using an empirical measure of the timescale

$$\tau \approx \langle \left\| S^2 \right\| \rangle^{-1} \tag{33}$$

In our work, we confirmed it was vital to performance to normalize the VGT

$$A' = \tau A \tag{34}$$

but found the network performance suffered considerably if the PH was normalized. One hypothesis is that this is an artifact of initialization of the network parameters. The PH should be normalized by τ^2, which for $Re = 240$ $\tau \approx 3e-3$. If one uses this normalization, the network does train, but it seems to start very far away from the optimal. It is possible this could remedied by scaling the initialization of the linear output layer as well as individualizing that layer's learning rate - this was not performed.

In calculation of the $g^{(i)}$ functions, we use a fully connected feed-forward network with 5 hidden layers of 50 nodes apiece with relu activation functions and a linear output layer. This network is trained using the ADAM optimizer with learning rate $5e-3$ decaying to a minimum of $1e-6$ via factors of 2 when the loss reaches a plateau.

4.2 Result reproduction

After implementing this network architecture and finding the hyperparameters above, the results in figure(3) were found. These results agree well with the original paper, and suggest the hyperparameter choices were accurate.

4.3 Temporal Convolution TBNN

Preliminary results suggest that the additional historical information of the VGT is in fact useful for the prediction of the PH. In figure(4), results of training the temporal convolution TBNN (convTBNN) are shown, improving both in the loss and eigenvector alignment metrics.

An interesting aspect of the convTBNN is the ability to inspect the kernel length - loosely the length over which the network found it useful to set weights away from zero. Figure(5) displays the weights as a function of kernel entry number. The sampling in time occurs every $20\Delta t$, where $\Delta t = 3e-4$ and the Kolmogorov timescale is $\tau \approx 3e-3$, so that the sampling occurs about every 2τ. We see that only the last three weights are nonzero, suggesting that a history of about 6τ is useful in prediction.

5 Conclusion

In this work (in progress), we reproduced the current state-of-the-art Lagrangian model to predict the PH, contextualized it with recent phenomenological advancements, and advanced the methodology by introducing and evaluating a new predictive model.

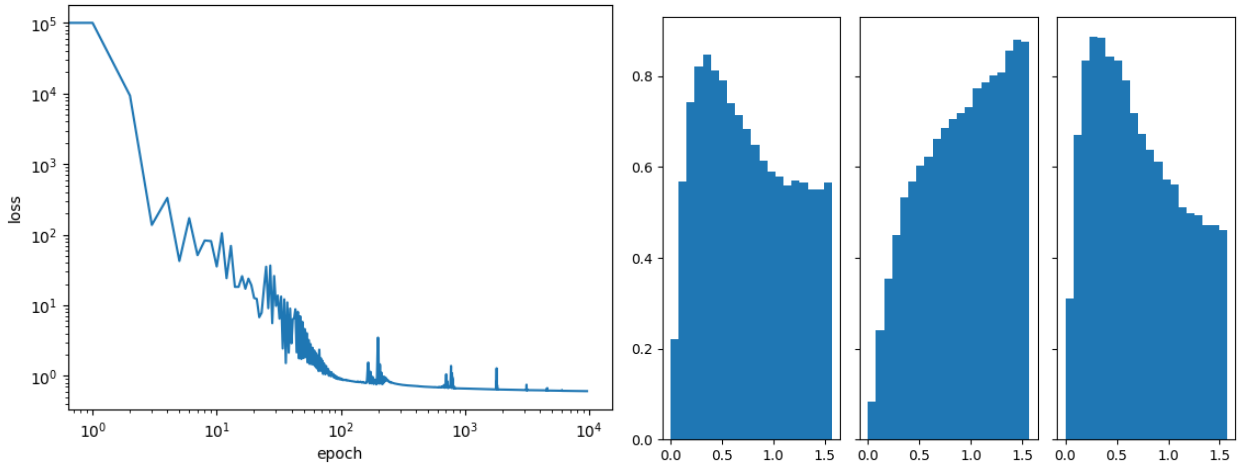

Figure 3: (left) Loss as a function of epoch for the TBNN, (right) PDFs of eigenvalue alignment for the pressure hessian, sorted from smallest to largest eigenvalue left to right. These plots agree well with results from Tian et.al's paper, suggesting the implementation is correct.

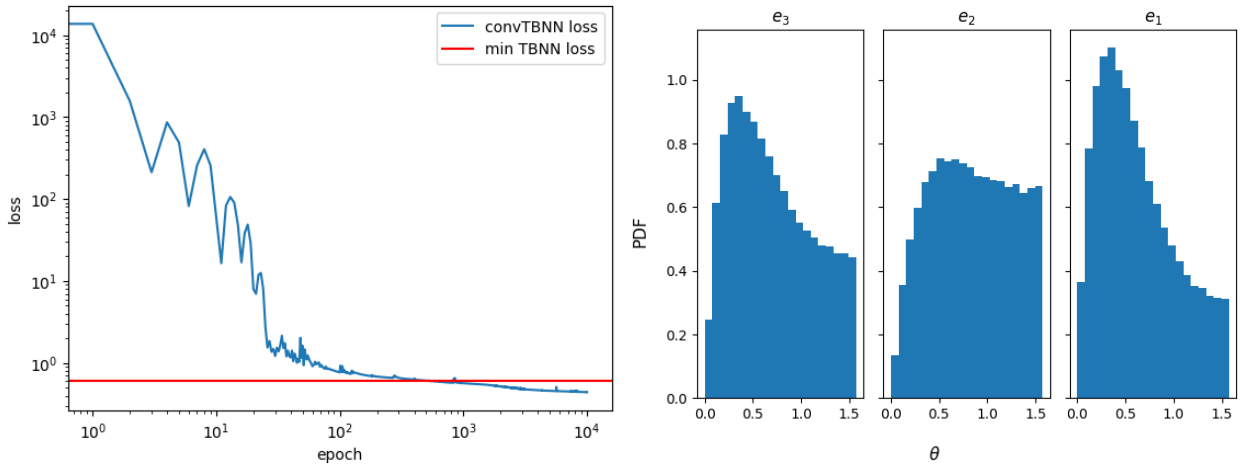

Figure 4: Preliminary results of training the convTBNN. (Left) shows the loss vs epoch, with the minimum loss obtained by the unmodified TBNN marked in horizontal red, and the convTBNN in blue. (Right) shows the alignment PDFs of eigenvectors of the PH. Here e_1 corresponds to the eigenvector with the greatest associated eigenvalue. Note that compared to fig(3), all alignment PDFs are better, shifting significant density from the misaligned (larger θ) to aligned.

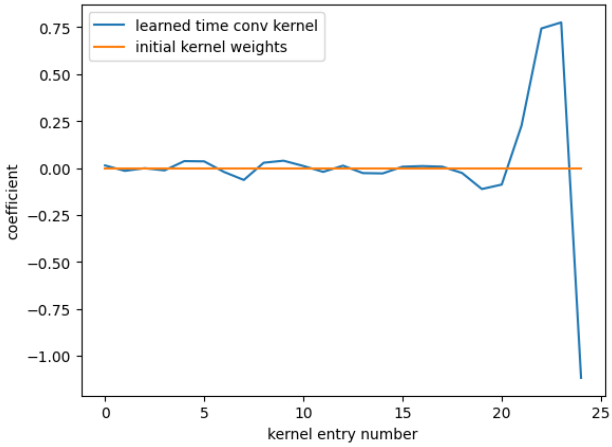

Figure 5: The weights of the learned temporal convolution kernel. This may act as a stand-in for a measure of significance for VGT history as it relates to the prediction of the PH. If these results hold across many trainings, it would suggest that only "recent" history is informative in prediction of the PH.

The TBNN is a powerful tool to capture symmetries, while allowing for the flexibility of a NN. We applied this model to homogeneous, isotropic turbulence to close the equations of the non-local pressure Hessian using only the local velocity gradient tensor. We contrasted this approach with a mathematically similar, but philosophically different approach using the RDGF model, motivated by the idea that the recent history of the VGT could better inform the prediction of the current PH.

By combining the two, we created the convTBNN, a first data-driven step towards biasing the statistics of the current PH prediction using the history of the VGT. We showed that we could outperform the local in time TBNN in two metrics, loss value and eigenvector alignment. We began to interpret the weights of the temporal convolution kernel, and related them to the hypotheses made in the phenomenological models (RFD, RDGF).

Further work needs to be done to ensure these results hold under evaluation of richer metrics such as the Q-R conditional mean tangents, and evaluate the accuracy and stability of the resulting differential equation for the VGT.

6 Appendix

6.1 Interpretability

Previous work by the authors [18] suggested there may be a latent space in the invariants when predicting the PH using the TBNN. The existence of a latent space could be important to the ability to interpret the results, e.g. searching for a functional form or providing physical insight; or if we believe in the statistical independence of the input (in our case there are many reasons to) it suggests the network struggles to exploit the additional information. In the case of [18] it seems to be the latter.

The proposed latent space is indeed persistent when training the network having normalized the PH using τ^2, in particular, the latent space claimed consists only of

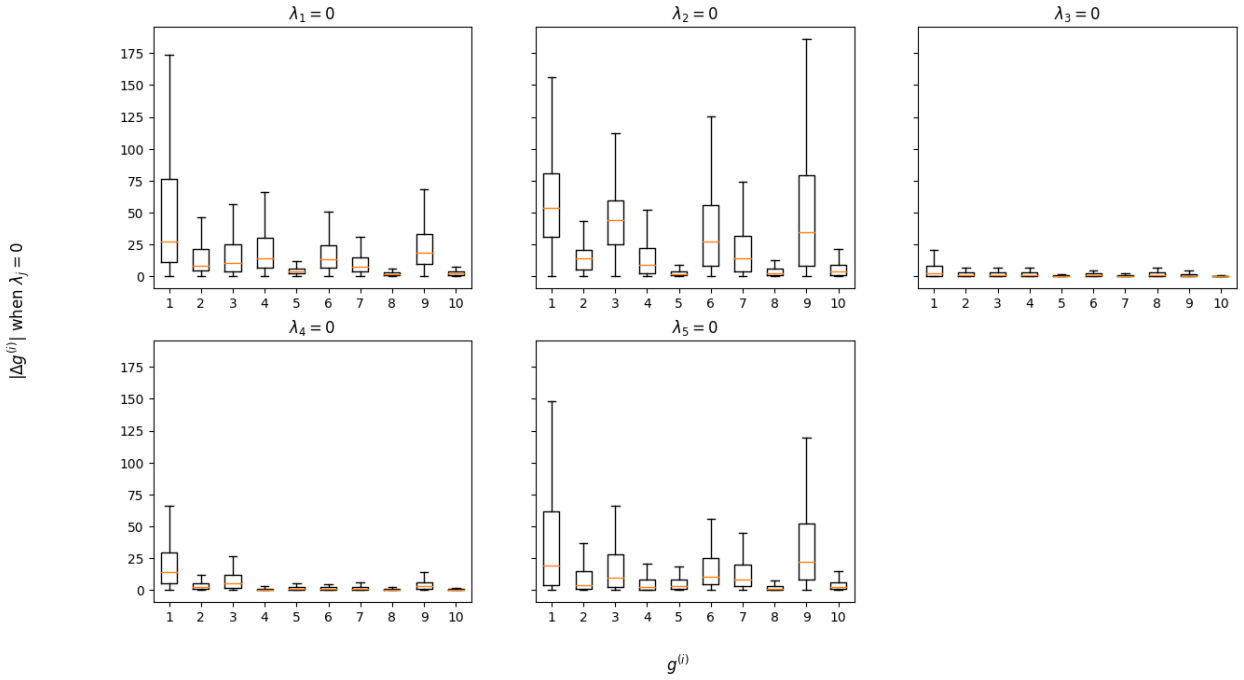

Figure 6: Distributions of sensitivity of $g^{(i)}$ to variability of λ_j. Previous work performed suggested that λ_{3-5} were unimportant to the prediction using the TBNN. This result was predicated upon the normalization of the PH using τ^2. While our plot reinforces that $\lambda_{3,4}$ do provide an order of magnitude smaller correction, the sensitivity of the network to λ_5 is of the same order of magnitude as that to $\lambda_{1,2}$. This suggests that by *not* normalizing the PH, we are able to exploit additional information in the invariants.

the low-order invariants. The hypothesis by the author was that the latent space emerged as a numerical artifact of strong normalization of the high-order invariants ($\lambda_{3,4} \to \tau^3\lambda_{3,4}$, and $\lambda_5 \to \tau^4\lambda_5$). But as we show in figure(6), there exists leading-order sensitivity to λ_5 (and not insignficant sensitivity to $\lambda_{3,4}$) when the PH is *not* normalized. This may indicate that the low-order latent space was a numerical artifact resulting from very small weights in the *output layer*, combined with strong normalization of high-order invariants.

To explore this question further, we applied an autoencoder to the set of invariants, as well as calculated the mutual information between the invariants.

The mutual information calculation, shown in table(1), indicates that while the invariants are not strictly independent, they are also far from highly correlated. The maximum value - biased away from the ideal value of 10 here by the k-nearest neighbors algorithm with $k = 5$, detailed in [35].

The autoencoder structure is a fully connected, feed forward neural network,

$$AE_\theta : \mathbb{R}^5 \to \mathbb{R}^5 \text{ via } \mathbb{R}^5 \to \mathbb{R}^4 \to \mathbb{R}^h \to \mathbb{R}^4 \to \mathbb{R}^5 \tag{35}$$

where h is the imposed latent dimension, illustrated in fig[7]. The nonlinear activation functions of AE_θ are the so-called "leaky-relu". We implement the "robust max-min" normalization, relying on the interquartile distance - namely, letting Q_1, Q_3 be the first

$I(\lambda_i, \lambda_j)$	λ_1	λ_2	λ_3	λ_4	λ_5
λ_1		0.26350727	1.40593169	0.40473788	0.72127464
λ_2			0.1968969	0.80102301	0.95679744
λ_3				0.43549834	0.49519878
λ_4					0.83990126
λ_5					

Table 1: A scaled mutual information $I(\lambda_i, \lambda_j)$, between invariants. Note the maximum value is near 10, the method used here introduces a bias however, namely via k-nearest neighbors with $k = 5$, as discussed in [35]

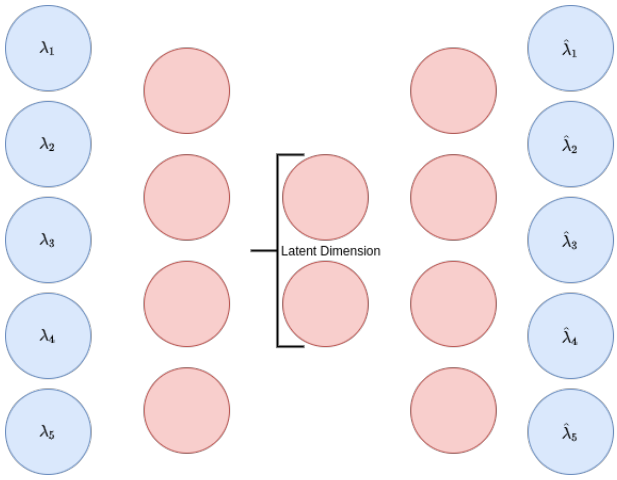

Figure 7: Autoencoder structure applied to the invariant input data.

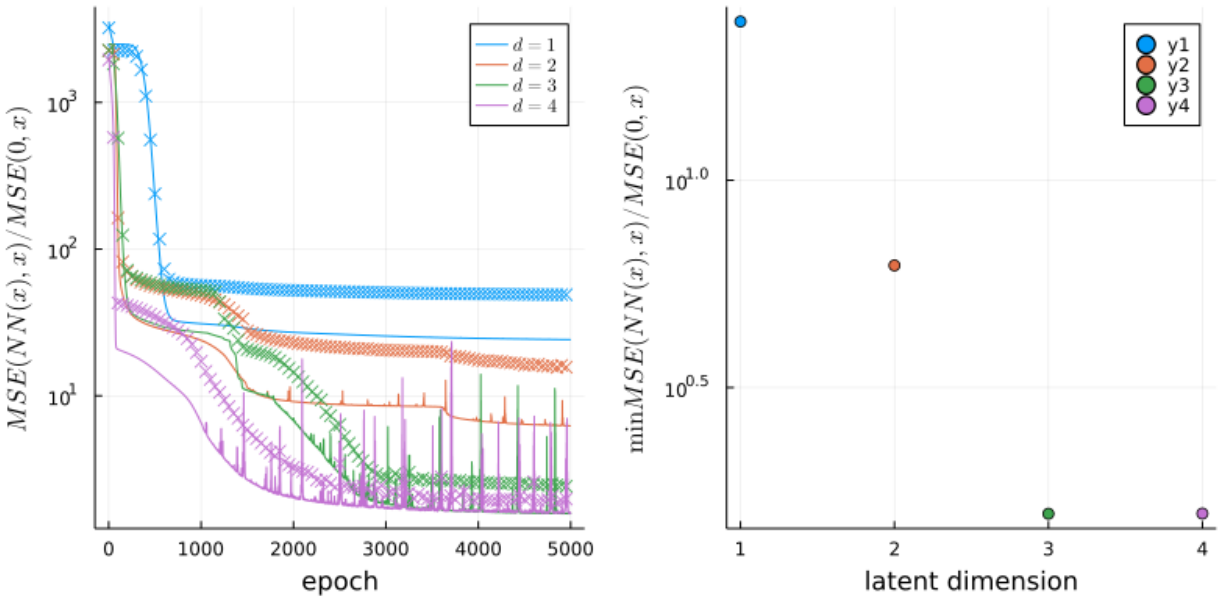

Figure 8: Autoencoder applied to the invariant input data, in an attempt for a nonlinear projection onto a latent space. (Left) shows the evolution of the loss according to the imposed latent dimension, where 'x's mark test loss, while solid lines mark training loss. (Right) shows the minimum loss as a function of latent dimension. Notice the two jumps at $h = 2$ and $h = 3$ - suggesting a sequence of increasing fidelity.

and third quartiles, we normalize our invariants as

$$\tilde{\lambda_i} = \frac{\lambda - Q_1}{Q_3 - Q_1} \tag{36}$$

This is intentionally different than the normalization used in the TBNN, as normalizing via timescale and optimizing using the mean-squared-error gives a very strong preference towards reconstructing only the low-order invariants. The robust max-min normalization is an attempt to set each on an equal footing, but itself is subject to biases introduced by differences in distribution shapes.

We use the ADAM optimizer with learning rate $\eta = 1e-3$, partitioning the dataset of 100k samples into 75% training, 25% test. The results are shown in fig[8] and may suggest a small reduction in dimension (i.e., $5 \to 3$). These results match the leading order contributions to the sensitivities of the $g^{(i)}$ as shown in fig(6), but we caution this suggestion as the low-order basis elements are still significantly influenced.

References

[1] Wm. T. Ashurst, A. R. Kerstein, R. M. Kerr, and C. H. Gibson. Alignment of vorticity and scalar gradient with strain rate in simulated navier–stokes turbulence. *The Physics of Fluids*, 30(8):2343–2353, 1987.

[2] R. Betchov. An inequality concerning the production of vorticity in isotropic turbulence. *Journal of Fluid Mechanics*, 1(5):497–504, 1956.

[3] Brian J. Cantwell. Exact solution of a restricted euler equation for the velocity gradient tensor. *Physics of Fluids A: Fluid Dynamics*, 4(4):782–793, 1992.

[4] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *CoRR*, abs/1806.07366, 2018.

[5] M. Chertkov, G. Falkovich, I. Kolokolov, and V. Lebedev. Normal and anomalous scaling of the fourth-order correlation function of a randomly advected passive scalar. *Phys. Rev. E*, 52:4924–4941, Nov 1995.

[6] Michael Chertkov, Alain Pumir, and Boris I. Shraiman. Lagrangian tetrad dynamics and the phenomenology of turbulence. *Physics of Fluids*, 11(8):2394–2410, 1999.

[7] L Chevillard, C Meneveau, L Biferale, and F Toschi. Modeling the pressure hessian and viscous laplacian in turbulence: comparisons with direct numerical simulation and implications on velocity gradient dynamics. *Physics of Fluids*, 20(10), 2008.

[8] Laurent Chevillard and Charles Meneveau. Lagrangian dynamics and statistical geometric structure of turbulence. *Physical review letters*, 97(17):174501, 2006.

[9] M. S. Chong, A. E. Perry, and B. J. Cantwell. A general classification of three-dimensional flow fields. *Physics of Fluids A: Fluid Dynamics*, 2(5):765–777, 1990.

[10] Mohammad Danish and Charles Meneveau. Multiscale analysis of the invariants of the velocity gradient tensor in isotropic turbulence. *Phys. Rev. Fluids*, 3:044604, Apr 2018.

[11] Rishita Das and Sharath S. Girimaji. On the reynolds number dependence of velocity-gradient structure and dynamics. *Journal of Fluid Mechanics*, 861:163–179, 2019.

[12] Rishita Das and Sharath S. Girimaji. Characterization of velocity-gradient dynamics in incompressible turbulence using local streamline geometry. *Journal of Fluid Mechanics*, 895:A5, 2020.

[13] D. A. Donzis, P. K. Yeung, and K. R. Sreenivasan. Dissipation and enstrophy in isotropic turbulence: Resolution effects and scaling in direct numerical simulations. *Physics of Fluids*, 20(4):045108, 2008.

[14] Adrienne L. Fairhall, Barak Galanti, Victor S. L'vov, and Itamar Procaccia. Direct numerical simulations of the kraichnan model: Scaling exponents and fusion rules. *Phys. Rev. Lett.*, 79:4166–4169, Nov 1997.

[15] G. Falkovich, K. Gawędzki, and M. Vergassola. Particles and fields in fluid turbulence. *Rev. Mod. Phys.*, 73:913–975, Nov 2001.

[16] Criston Hyett, Michael Chertkov, Yifeng Tian, and Daniel Livescu. Machine learning statistical lagrangian geometry of turbulence. In *APS Division of Fluid Dynamics Meeting Abstracts*, pages S01–024, 2020.

[17] Criston Hyett, Michael Chertkov, Yifeng Tian, Daniel Livescu, and Mikhail Stepanov. Machine learning statistical evolution of the coarse-grained velocity gradient tensor. In *APS Division of Fluid Dynamics Meeting Abstracts*, pages E31–009, 2021.

[18] Criston Hyett, Yifeng Tian, Michael Woodward, Michael Chertkov, Daniel Livescu, and Mikhail Stepanov. Applicability of machine learning methodologies

to model the statistical evolution of the coarse-grained velocity gradient tensor. *Bulletin of the American Physical Society*, 2022.

[19] Michael Innes, Elliot Saba, Keno Fischer, Dhairya Gandhi, Marco Concetto Rudilosso, Neethu Mariya Joy, Tejan Karmali, Avik Pal, and Viral Shah. Fashionable modelling with flux. *CoRR*, abs/1811.01457, 2018.

[20] Mike Innes. Flux: Elegant machine learning with julia. *Journal of Open Source Software*, 2018.

[21] Perry L. Johnson and Charles Meneveau. Large-deviation joint statistics of the finite-time lyapunov spectrum in isotropic turbulence. *Physics of Fluids*, 27(8):085110, 2015.

[22] Perry L. Johnson and Charles Meneveau. A closure for lagrangian velocity gradient evolution in turbulence using recent-deformation mapping of initially gaussian fields. *Journal of Fluid Mechanics*, 804:387–419, 2016.

[23] Perry L. Johnson and Charles Meneveau. Turbulence intermittency in a multiple-time-scale navier-stokes-based reduced model. *Phys. Rev. Fluids*, 2:072601, Jul 2017.

[24] Robert H. Kraichnan. Convection of a passive scalar by a quasi-uniform random straining field. *Journal of Fluid Mechanics*, 64(4):737–762, 1974.

[25] J. M. Lawson and J. R. Dawson. On velocity gradient dynamics and turbulent structure. *Journal of Fluid Mechanics*, 780:60–98, 2015.

[26] Thomas S. Lund and Michael M. Rogers. An improved measure of strain state probability in turbulent flows. *Physics of Fluids*, 6(5):1838–1847, 1994.

[27] Charles Meneveau. Lagrangian dynamics and models of the velocity gradient tensor in turbulent flows. *Annual Review of Fluid Mechanics*, 43:219–245, 2011.

[28] Parviz Moin and Krishnan Mahesh. Direct numerical simulation: A tool in turbulence research. *Annual Review of Fluid Mechanics*, 30(1):539–578, 1998.

[29] Koji Ohkitani and Seigo Kishiba. Nonlocal nature of vortex stretching in an inviscid fluid. *Physics of Fluids*, 7(2):411–421, 1995.

[30] ANDREW OOI, JESUS MARTIN, JULIO SORIA, and M. S. CHONG. A study of the evolution and characteristics of the invariants of the velocity-gradient tensor in isotropic turbulence. *Journal of Fluid Mechanics*, 381:141–174, 1999.

[31] S. B. Pope. A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, 72(2):331–340, 1975.

[32] Christopher Rackauckas and Qing Nie. Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in julia. *The Journal of Open Research Software*, 5(1), 2017. Exported from https://app.dimensions.ai on 2019/05/05.

[33] Maziar Raissi and George E. Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *CoRR*, abs/1708.00588, 2017.

[34] Hannes Risken and Frank Till. *The Fokker-Planck Equation*. Springer,Berlin,Heidelberg, 1996.

[35] Brian C Ross. Mutual information between discrete and continuous data sets. *PloS one*, 9(2):e87357, 2014.

[36] A. J. M. Spencer and R. S. Rivlin. Further results in the theory of matrix polynomials. *Arch. Rational Mech. Anal.*, 4:214–230, 1959.

[37] Anthony James Merrill Spencer and Ronald S Rivlin. The theory of matrix polynomials and its application to the mechanics of isotropic continua. *Archive for rational mechanics and analysis*, 2:309–336, 1958.

[38] K. R. Sreenivasan and R. A. Antonia. The phenomenology of small-scale turbulence. *Annual Review of Fluid Mechanics*, 29(1):435–472, 1997.

[39] Hendrik Tennekes and John Leask Lumley. *A first course in turbulence*. MIT press, 1972.

[40] Yifeng Tian, Yen Ting Lin, Marian Anghel, and Daniel Livescu. Data-driven learning of mori–zwanzig operators for isotropic turbulence. *Physics of Fluids*, 33(12):125118, 2021.

[41] Yifeng Tian, Daniel Livescu, and Michael Chertkov. Physics-informed machine learning of the lagrangian dynamics of velocity gradient tensor. *Physical Review Fluids*, 6(9):094607, 2021.

[42] P. Vieillefosse. Internal motion of a small element of fluid in an inviscid flow. *Physica A: Statistical Mechanics and its Applications*, 125(1):150–162, 1984.

[43] M Wan, S Chen, G Eyink, C Meneveau, E Perlman, R Burns, Y Li, A Szalay, and S Hamilton. Johns hopkins turbulence database (jhtdb), 2016.

[44] Q.-S. Zheng. On the representations for isotropic vector-valued, symmetric tensor-valued and skew-symmetric tensor-valued functions. *International Journal of Engineering Science*, 31(7):1013–1024, 1993.