

STRUCTURE PRESERVATION FOR THE DEEP NEURAL NETWORK MULTIGRID SOLVER*

NILS MARGENBERG[†], CHRISTIAN LESSIG[‡], AND THOMAS RICHTER[§]

Abstract. The simulation of partial differential equations is a central subject of numerical analysis and an indispensable tool in science, engineering, and related fields. Existing approaches, such as finite elements, provide (highly) efficient tools but deep neural network-based techniques emerged in the last few years as an alternative with very promising results. We investigate the combination of both approaches for the approximation of the Navier-Stokes equations and to what extent structural properties such as divergence freedom can and should be respected. Our work is based on DNN-MG, a deep neural network multigrid technique, that we introduced recently and which uses a neural network to represent fine grid fluctuations not resolved by a geometric multigrid finite element solver. Although DNN-MG provides solutions with very good accuracy and is computationally highly efficient, we noticed that the neural network-based corrections substantially violate the divergence freedom of the velocity vector field. In this contribution, we discuss these findings and analyze three approaches to address the problem: a penalty term to encourage divergence freedom of the network output; a penalty term for the corrected velocity field; and a network that learns the stream function and which hence yields divergence-free corrections by construction. Our experimental results show that the third approach based on the stream function outperforms the other two and not only improves the divergence freedom but also the overall fidelity of the simulation.

Key words. multigrid methods, incompressible Navier-Stokes equations, machine learning, artificial neural networks

AMS subject classifications. 65M55, 76D05

1. Introduction. The tremendous progress on deep learning in the last decade for applications such as image recognition and machine translation [28] has motivated their use also for the simulation of physical systems. Despite these promising results that have already been obtained, many fundamental questions remain open. One of them is how physical constraints, such as energy and momentum conservation or divergence restriction, can be integrated into a neural network-based simulation—and if an explicit consideration of this is necessary or meaningful.

In the literature, two principle approaches to enforce physical constraints have been proposed. In the first one, they are imposed by construction, e.g., through an appropriate network design or by representing the physical quantities so that their properties and relationship have to be satisfied. Examples are symplectic neural network [12, 38] that impose symplecticity by construction or the representation of fluid velocity fields in a divergence-free basis with learning being performed on the basis function coefficients [49, 54]. The second and more common approach to physical constraints in neural networks is to encourage that these hold but not strictly enforce them. Such a “soft” formulation is typically accomplished using a penalty term in the loss function but can also be realized, e.g., with a Bayesian formulation with the physical constraints (including an entire partial differential equation) as priors. The latter approach also naturally accommodates real-world data where properties such as energy conservation almost never hold perfectly; e.g., because a system is not closed. With a penalty term it is, however, difficult to control to what extent a constraint is satisfied after training and its addition to the loss function increases the training complexity and can hinder convergence.

*Received December 15, 2020. Accepted April 26, 2021. Published online on December 21, 2021. Recommended by Peter Benner.

[†]Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany (margenbn@hsu-hh.de).

[‡]University of Magdeburg, Institute for Simulation and Graphics, Universitätsplatz 2, 39104 Magdeburg, Germany (christian.lessig@ovgu.de).

[§]University of Magdeburg, Institute for Analysis and Numerics, Universitätsplatz 2, 39104 Magdeburg, Germany (thomas.richter@ovgu.de).

In this work, we explore how divergence restrictions of the fluid velocity vector field can be ensured when a neural network is used in the simulation of the incompressible Navier-Stokes equations. In particular, we build on the recently introduced Deep Neural Network Multigrid Solver (DNN-MG) [36] that tightly integrates a neural network into a multigrid finite element solver to replace the computations on one or multiple finest mesh levels by a neural network-based prediction. In [36], the efficiency of DNN-MG for the simulation of the incompressible, instationary Navier-Stokes equations was demonstrated. However, the predicted velocity vector field is not divergence free. We hence study in this paper three different approaches to encourage or enforce divergence freedom within the DNN-multigrid formulation. First, we modify the network’s loss function and include a penalty term to limit the divergence of the predicted velocity field. Second, we penalize the divergence of the corrected velocity field, which has the potential to also reduce the divergence residual of the finite element velocity field. And third, we design a neural network that learns the stream function, i.e., the potential associated with the velocity vector field, and hence provides by construction divergence-free corrections.

Our results demonstrate that encouraging or enforcing divergence freedom improves the overall quality of the simulations, measured, e.g., by the drag and lift functionals. The three approaches we developed lead, however, only to rather modest reductions in the divergence. For the two approaches based on a penalty term, this results from a delicate balance between prediction accuracy and divergence freedom that needs to be attained but that it is difficult to satisfy in practice. The approach based on the stream function provides overall the best results and achieves lift and drag functionals close to a fine reference solution. At the same time, the corresponding network is more efficient to train.

The remainder of the paper is structured as follows. After discussing related work, we recall the DNN-MG solver in Section 3. In Section 4 we describe our approaches to ensure divergence freedom within DNN-MG and present and discuss our numerical results. We conclude in Section 5 where we also provide an outlook for possible directions for future work.

2. Related work. Recently, there has been an increasing interest to employ (deep) neural networks for the simulation of physical systems. In the following, we will discuss existing approaches with an emphasis on those that consider partial differential equations and that are most pertinent to our own work.

Neural network-based techniques for partial differential equations can, broadly speaking, be classified into three categories. The first one, e.g., the DeepRitz method by E and Yu [16] or the physics-informed neural networks by Raissi, Karniadakis, and collaborators [34, 43], use neural networks in place of a basis or frame to represent solutions. Training then amounts to fitting the network representation to the initial or boundary value problem. This has hence to be repeated for each problem instance, but the approach yields typically a mesh-free solution that scales well to high dimensions [16]. The methodology also applies to a wide range of partial differential equations. The second approach takes inspiration from classical time stepping schemes and considers the solution of the PDE as the problem of predicting the next state from the previous ones. The well developed network architectures for sequential data can hence be employed for this approach, e.g., recurrent neural networks (RNN) with Long-Short-Term-Memory units (LSTM) [24] or Gated Recurrent Units (GRUs) [13], or temporal convolutional networks; cf. [1]. Examples of this approach are [25, 51, 55] and the work by Han, Jentzen, and E [22], who use a formulation of semi-linear, parabolic PDEs as backward stochastic differential equations and then realize each time step of these as a layer in a deep neural network. Since we combine a classical finite element based simulation, which relies on a time stepping scheme, with a neural network, our work also falls into this second

category. The third class of neural network-based techniques for partial differential equations uses these to represent solution operators. They hence use the initial condition as well as parameters on which the solution depends as input to the neural network and it then maps these to the solution. In [7], for example, the solution operator for parametric partial differential equations is learned based on a discretization obtained via model reduction and [29, 30] learn mappings between infinite dimensional spaces that can be solution operators. Lu, Jin, and Karniadakis [33] learn nonlinear operators from data, based on the result that any such operator can be approximated by a neural network [11].

For flow problems, Raissi et al. [44] demonstrated that the velocity and pressure of the Navier-Stokes equations can be learned to good approximation only from observing passively advected tracer particles. Nabian and Meidani [40], Yang and Perdikaris [56], and Raissi and Karniadakis [42] exploited that not only observations are available but also a known analytic model. These authors hence also use the error between the network prediction and the model as an additional penalty term in the training loss. Kasim et al. [27] recently demonstrated neural network-based simulations for a broad range of applications, including fluid dynamics, by also optimizing the network architecture itself during training. Eichinger, Heinlein, and Klawonn [35] use convolutional neural networks and techniques from image processing to learn flow patterns for the Navier-Stokes flow around objects of different shape.

Next to the aforementioned direct approaches to the simulation of partial differential equations using neural networks, there have been different attempts to integrate these into existing numerical formulations. As in our work, the objective is then to combine the benefits of neural networks and classical methods and obtain techniques that would be difficult or impossible with either approach alone. For elasticity, neural network-based representations for constitutive relations were learned in [6, 46, 52] and then used in classical simulations, e.g., based on finite elements. Wiewel et al. [55] presented a simulation of the Navier-Stokes equation where an LSTM-based deep neural network is used to predict the pressure correction within a classical finite element simulation. In [32] a convolutional neural network is interpreted as a finite difference discretization of a PDE. This was also proposed in [47]. Bar-Sinai, Hoyer, Hickeney, and Brenner [2] learn general sub-grid scale models, used, e.g., with finite difference or finite volume discretization. In the context of large eddy-type simulations, the learning of closure models has also been considered, e.g., in [31]. Wan, Dodov, Lessig, Dijkstra, and Sapsis [54] combine a wavelet-based discretization and perform learning in the coefficient space where space-frequency correlations are resolved but, so far, without simulation. Recently, Stevens, and Colonius [51] enriched finite difference and finite volume simulations with a deep neural network to resolve sharp discontinuities (as in Burger's equations) and to deal with PDEs with chaotic behavior.

To ensure that neural network-based predictions respect physical invariants, such as energy conservation or divergence freedom, network architectures tailored towards physical simulations have been proposed. Next to an implicit enforcement as in Physics Informed Neural Network (PINN) [34], various authors proposed architecture where desired properties hold by construction. Finizi, Wang, and Wilson [18] and Cranmer et al. [15] learn the dynamics of classical mechanical systems by directly learning the Hamiltonian or Lagrangian functions. In related work, [12, 26, 38] developed neural networks that incorporate the symplectic structure of Hamiltonian mechanics and demonstrated that this improves generalization and prediction accuracy. Based on the work by Bruna and Mallat [10], Bietti and Mairal [8] proposed neural network architectures that are invariant under group transformation, including diffeomorphisms. A neural network that is directly defined on a Lie group, and hence also enjoys group equivariance, was proposed by Bekkers [5].

3. Deep Neural Network Multigrid Solver. In this section we summarize the Deep Neural Network Multigrid Solver (DNN-MG) introduced in [36]. For a detailed description the reader is referred to the original paper.

3.1. Finite element discretization of the incompressible Navier-Stokes equations. We want to solve the incompressible, stationary Navier-Stokes equations given by

$$(3.1) \quad \begin{aligned} \partial_t v + (v \cdot \nabla)v - \frac{1}{\text{Re}} \Delta v + \nabla p &= f \quad \text{on } [0, T] \times \Omega \\ \nabla \cdot v &= 0 \quad \text{on } [0, T] \times \Omega, \end{aligned}$$

where $v: [0, T] \times \Omega \rightarrow \mathbb{R}^2$ is the velocity, $p: [0, T] \times \Omega \rightarrow \mathbb{R}$ the pressure, $\text{Re} > 0$ the Reynolds number, and f an external force. The initial and boundary conditions are given by

$$(3.2) \quad \begin{aligned} v(0, \cdot) &= v_0(\cdot) \quad \text{on } \Omega \\ v &= v^D \quad \text{on } [0, T] \times \Gamma^D \\ \frac{1}{\text{Re}}(\vec{n} \cdot \nabla)v - p\vec{n} &= 0 \quad \text{in } [0, T] \times \Gamma^N, \end{aligned}$$

where \vec{n} denotes the outward facing unit normal on the boundary $\partial\Omega$ of the domain. On the outflow boundary Γ^N we consider the do-nothing outflow condition [23], which is well established to model artificial boundaries; see the discussion in [9] on possible shortcomings and a variation that is able to guarantee uniqueness of a solution.

We discretize (3.1) using a weak finite element formulation with $v_h, \phi_h \in V_h = [W_h^{(2)}]^d$ and $p_h, \xi_h \in L_h = W_h^{(2)}$, where ϕ_h and ξ_h are test functions and $W_h^{(r)}$ is the space of continuous functions that are polynomials of degree r on each mesh element $T \in \Omega_h$ in the mesh domain Ω_h . The resulting equal order finite element pair $V_h \times L_h$ does not fulfill the inf-sup condition. We hence use stabilization terms of local projection type [3] with parameter $\alpha_T = \alpha_0 \cdot \text{Re} \cdot h_T^2$ and projection $\pi_h: W_h^{(2)} \rightarrow W_h^{(1)}$ into the space of linear polynomials.

With the second order Crank-Nicolson method for time discretization, the solution of (3.1) subject to (3.2) determines at each time step n the state $x_n = (v_n^1, \dots, v_n^d, p_n)$ such that

$$(3.3) \quad \mathcal{A}_n(x_n) = f_n,$$

where

$$(3.4) \quad \begin{aligned} [\mathcal{A}_n(x_n)]_i &:= (\nabla \cdot v_n, \xi_h^i) + \sum_{T \in \Omega_h} \alpha_T (\nabla(p_n - \pi_h p_n), \nabla(\xi_h^i - \pi_h \xi_h^i)), \\ [f_n]_i &:= 0 \end{aligned}$$

for all pressure unknowns $i = 1, \dots, N_p$ and

$$\begin{aligned} [\mathcal{A}_n(x_n)]_{i+N_p} &:= \frac{1}{k}(v_n, \phi_h^i) + \frac{1}{2}(v_n \cdot \nabla v_n, \phi_h^i) + \frac{1}{2\text{Re}}(\nabla v_n, \nabla \phi_h^i) - (p_n, \nabla \cdot \phi_h^i) \\ [f_n]_{i+N_p} &:= \frac{1}{k}(v_{n-1}, \phi_h^i) + \frac{1}{2}(f_n, \phi_h^i) + \frac{1}{2}(f_{n-1}, \phi_h^i) \\ &\quad - \frac{1}{2}(v_{n-1} \cdot \nabla v_{n-1}, \phi_h^i) - \frac{1}{2\text{Re}}(\nabla v_{n-1}, \nabla \phi_h^i), \end{aligned}$$

for all velocity unknowns $i = 1, \dots, N_v$ with corresponding test functions ϕ_h^i and ξ_h^i . By $k > 0$ we denote the time step size. Equation (3.3) is a large nonlinear system of algebraic equations that is solved by Newton's method based on the initial guess $x_n^{(0)} = (v_{n-1}, p_{n-1})$ and using the iteration

$$(3.5) \quad \mathcal{A}'_n(x_n^{(l-1)}) w^{(l)} = f_n - \mathcal{A}_n(x_n^{(l-1)}), \quad x_n^{(l)} = x_n^{(l-1)} + w^{(l)}$$

Algorithm 1 DNN-MG for the solution of the Navier-Stokes equations. Lines 6-9 (blue) provide the modifications of the DNN-MG method compared to a classical Newton-Krylov simulation with geometric multigrid preconditioning.

```

1: for all time steps  $n$  do
2:   while not converged do ▷ Newton's method in (3.5)
3:      $\delta x_i \leftarrow \text{MULTIGRID}(L, A_L^n, b_L^n, \delta x_i)$  ▷ Geometric multigrid with  $x = (p_n^L, v_n^L)$ 
4:      $x_{i+1} \leftarrow x_i + \epsilon \delta x_i$ 
5:   end while
6:    $(\tilde{p}_n^{L+1}, \tilde{v}_n^{L+1}) \leftarrow \mathcal{P}(p_n^L, x_n^L)$  ▷ Prolongation on level  $L + 1$ 
7:    $r_n^{L+1} \leftarrow f_n^{L+1} - \mathcal{A}_n(\tilde{p}_n^{L+1}, \tilde{v}_n^{L+1})$  ▷ Computation of the residual
8:    $d_n^{L+1} \leftarrow \mathcal{N}(\tilde{r}_n^{L+1}, \tilde{v}_n^{L+1}, \Omega_L, \Omega_{L+1})$  ▷ Prediction of velocity correction
9:    $b_{n+1}^{L+1} \leftarrow \text{Rhs}(\tilde{v}_n^{L+1} + d_n^{L+1}, f_n, f_{n+1})$  ▷ Set up rhs of (3.3) for next time step
10:   $b_{n+1}^L \leftarrow \mathcal{R}(b_{n+1}^{L+1})$  ▷ Restriction of rhs to level  $L$ 
11: end for

```

for $l = 1, 2, \dots$. Here we denote by $\mathcal{A}'(x^{(l-1)})$ the Jacobian of \mathcal{A} at $x^{(l-1)}$, which for the problem at hand can be computed analytically; cf. [45, Section 4.4.2].

Each Newton step requires the solution of the linear system in (3.5), where the system matrix $\mathcal{A}'(x^{(l-1)})$ is sparse but non-symmetric and not definite due to the saddle point structure of the underlying Navier-Stokes equations. To approximate the solution to (3.5) with optimal robustness, we employ the generalized minimal residual method (GMRES) introduced by Saad [48]. The convergence is accelerated through a preconditioner that is realized by a single sweep of a geometric multigrid solver.

The geometric multigrid method is based on a hierarchical approximation of a linear system on a sequence of finite element spaces $V_0 \subset V_1 \subset \dots \subset V_L$ defined over a hierarchy of meshes $\Omega_0, \dots, \Omega_L = \Omega_h$. Instead of treating the linear system on its domain, given by the finest mesh level L (as one would do in a traditional solver) the high frequencies there are smoothed and the remaining errors are treated on lower levels. This is accomplished by computing the residual on level L and restricting it to the next coarse level $L - 1$. There, the process is repeated and this is performed recursively until the coarsest level has been reached where a direct solver is employed. The updates from the coarser levels are then prolonged back to level L and the coarse-to-fine and fine-to-coarse iteration is repeated until a prescribed error tolerance has been met. The mesh transfer from fine to coarse is accomplished with L^2 -projections, known as restrictions, and from coarse to fine with interpolations, known as prolongations. The smoothing that preceded the restrictions is realized with a simple smoothing operator $S(A_l, b_l, x_l)$ that yields an approximate solution of the linear system $A_l x_l = b_l$, i.e., $S(A_l, b_l, x_k) \approx A_l^{-1} b_l$ and that aims to quickly reduce all high frequency components of the residual $b_l - A_l x_l$. In our implementation we use a simple iteration of Vanka-type [53], which allows for easy parallelization and gives very good performance with less than 5 pre- and post-smoothing steps [17].

3.2. Deep Neural Network Multigrid Solver. The solution of the Newton iteration in (3.5) using GMRES and the geometric multi-grid method is highly efficient. However, since one has multiple GMRES steps for every step of the Newton method and for each one also one up- and down-sweep of the multigrid method, there is still a substantial amount of computations. The majority is thereby required for the finest mesh level (typically about three quarters of the computations for the multigrid method) so that a refinement of the mesh,

or, equivalently, the use of an additional mesh level $L + 1$, incurs a substantial increase in computational costs. The Deep Neural Network Multigrid Solver (DNN-MG) addresses this bottleneck and replaces computations on one or multiple finest mesh levels with a neural network-based prediction. An overview of DNN-MG is provided in Algorithm 1.

To ensure that the DNN-MG solver is efficient and flexible, its neural network is build around three design principles:

1. The neural network operates patch-wise, i.e., locally on small neighborhoods of the mesh domain Ω_{L+k} . This ensures the generalizability of DNN-MG to different flow regimes and meshes.
2. The neural network has memory. This ensures that complex flow behavior can be predicted and that corrections are coherent in time.
3. The neural network uses the residual of the nonlinear problem in (3.3) on level $L + k$ as input. This provides rich information about the required correction.

Together, these design principles allows us to use to a surprisingly compact neural network architecture that at the same time generalizes well to different flow regimes and domains. In particular, the implementation in [36] has just 8634 trainable parameters with an architecture based on Gated Recurrent Units. It can hence be trained in a few hours with a small number of example flows as training set.

As shown in Algorithm 1, the neural network-based correction of DNN-MG is applied at the end of every time step after the Newton solver has computed an updated velocity v_n^L on level L ; cf. (3.5). To obtain the correction, v_n^L is first prolonged to level $L + 1$, yielding $\tilde{v}_n^{L+1} = \mathcal{P}(v_n^L)$. The neural network is then employed individually for each patch P_i (a mesh element on level L in the most simple case as in [36]) and predicts a velocity correction $d_{n,i}^{L+1}$. The inputs to the neural network are thereby also entirely local and include the residual of (3.3) on level $L + 1$, the prolonged velocity \tilde{v}_n^{L+1} , geometric properties such as the patch's aspect ratio, and possibly other quantities such as the Péclet number over the patch. With the correction d_n^{L+1} , a provisional right hand side $b_{n+1}^{L+1} = \text{Rhs}(\tilde{v}_n^{L+1} + d_n^{L+1}, f_n, f_{n+1})$ of (3.5) is formed on level $L + 1$ and then restricted to level L , i.e., we compute $b_{n+1}^L = \mathcal{R}(b_{n+1}^{L+1})$. Restricting the right hand side instead of the corrected velocity $\tilde{v}_n^{L+1} + d_n^{L+1}$ is thereby essential to propagate information back to level L since, ideally, $\mathcal{R}(\mathcal{P}(v_n^L) + d_n^{L+1}) = v_n^L$. The corrected b_{n+1}^L is then used in the next time step in the Newton solve, which is again improved by a neural network-based correction at the end of the time step. Note that DNN-MG only corrects the velocity but not the pressure. The latter enters only indirectly through the network input in form of the fine mesh residual that is computed in step 7 of the algorithm. In Section 4.2.2 we will discuss a modification of DNN-MG that will learn velocity updates in a stream function formulation. These corrections are strictly divergence free and hence implicitly also respect the pressure space.

The neural network of DNN-MG is trained using a high fidelity finite element solution obtained on the fine mesh level $L + 1$ and with the loss function

$$(3.6) \quad \mathcal{L}(v^L, v^{L+1}; d^{L+1}) := \sum_{n=1}^N \sum_{P_i \in \Omega_{L+1}} \|v_{n+1}^{L+1} - (v_n^L + d_n^{L+1})\|_{l^2(P_i)}^2.$$

Here, N is the number of time-steps that make up the training data and, since DNN-MG operates strictly local over patches, for each one the loss accumulates the local residuals in the second sum. Due to this strictly local structure of DNN-MG, a very large training data set is generated by a single resolved simulation.

In [36] it was demonstrated that the DNN-MG algorithm substantially improves lift and drag functionals for the classical channel flow around a cylinder and generalized well to flow

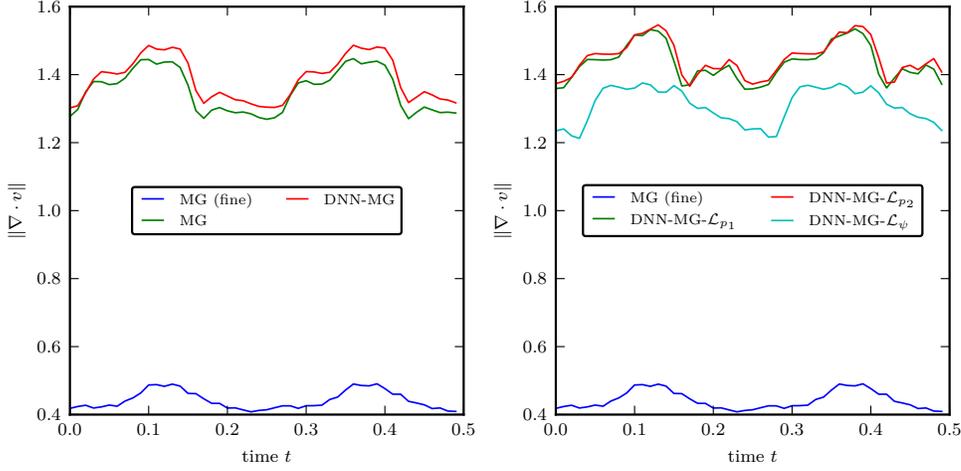


FIG. 4.1. Left: Divergence $\|\nabla \cdot v\|$ for the coarse mesh (MG), the high fidelity (MG fine) and the deep neural network multigrid solution (DNN-MG). Right: Also shown are the results for DNN-MG with the divergence of the DNN’s output as penalty term (DNN-MG- \mathcal{L}_{p_1}), DNN-MG with the divergence of the corrected velocity field as penalty term (DNN-MG- \mathcal{L}_{p_2}) and enforcing divergence freedom by construction (DNN-MG- \mathcal{L}_{ψ}) using the stream function formulation.

regimes not seen during training, e.g., a flow without an obstacle or a flow in an L-shaped domain.

4. Enforcing divergence freedom in DNN-MG. In our first study [36], we have shown that DNN-MG is able to increase the approximation quality at very little extra cost. The error in the velocity vector field but also in derived quantities such as the drag and lift functionals of the flow around an obstacle are reduced considerably with an increase in the computation time that is orders of magnitude below those for adding an additional mesh level. We have, however, also observed that the network is not good at ensuring a solenoidal velocity vector field. Although the loss-function aims at minimizing the distance to the high fidelity solution, which weakly satisfies divergence freedom to a certain degree, the neural network minimizes (3.6) with a vector field with a substantial divergence.

4.1. Violation of divergence freedom in existing formulation. The quadratic equal-order finite element approach $(v_h, p_h) \in V_h \times L_h$ that provides the basis for the DNN-MG algorithm in our current implementation is stabilized with local projections [3]; cf. (3.4). The discrete divergence equation thus reads

$$(\nabla \cdot v_n, \xi_h) + \sum_{T \in \Omega_h} \alpha_T (\nabla(p_n - \pi_h p_n), \nabla(\xi_h - \pi_h \xi_h)) = 0 \quad \forall \xi_h \in L_h,$$

and since $\nabla \cdot V_h \not\subset L_h$ only a perturbed divergence condition is satisfied

$$(4.1) \quad \|\nabla \cdot v_n\|_{L^2(\Omega)} = \|\nabla \cdot (v - v_n)\|_{L^2(\Omega)} \leq c \|v - v_n\|_{H^1(\Omega)}.$$

Figure 4.1, left, displays the total divergence $\|\nabla \cdot v(t)\|_{L^2(\Omega)}$ over the simulation domain for the coarse mesh solution (MG), the high fidelity solution (MG fine) and the DNN-MG solver for the experiments described in Section 4.3. While the fine mesh solution reduces the divergence at the expected rate, DNN-MG even further disturbs the divergence instead of improving it.

4.2. Approaches for enforcing divergence freedom. In the following we will discuss three approaches for enforcing divergence freedom. Numerical results will be presented afterwards in Section 4.3.

4.2.1. A penalty approach for divergence-free learning. The first approach to enforce divergence freedom of the neural network output of the DNN-MG method is through a modification of the loss function by a squared divergence as penalty term. One then has

$$(4.2) \quad \mathcal{L}_{p_1}(v^L, v^{L+1}; d^{L+1}) := \mathcal{L}(v^L, v^{L+1}; d^{L+1}) + \gamma \sum_{n=1}^N \sum_{P_i \in \Omega_h} \|\nabla \cdot d^{L+1}\|_{L^2(P_i)}^2,$$

where $\mathcal{L}(v^L, v^{L+1}; d^{L+1})$ is the original loss function in (3.6) and $d^{L+1} = \mathcal{N}(v^L, \tilde{v}^{L+1})$ is the correction by the network based on the coarse solution v_n^L and its prolongation \tilde{v}^{L+1} . The parameter $\gamma > 0$ is used to control the weighting of the loss function. We discuss its choice in Section 4.3. Equation (4.2) penalizes the divergence of the correction and, through this, the neural network should at least not further distort the divergence of the velocity field.

An obvious extension of this idea is to penalize the divergence of the corrected solution $v^L + d^{L+1}$. This is, in principle, advantageous since it could also correct for the divergence residual in the finite element formulation; cf. (4.1) and Figure 4.1, left. We thus introduce the second, modified loss function

$$(4.3) \quad \mathcal{L}_{p_2}(v^L, v^{L+1}; d^{L+1}) := l(v^L, v^{L+1}; d^{L+1}) + \gamma \sum_{n=1}^N \sum_{P_i \in \Omega_h} \|\nabla \cdot (v^L + d^{L+1})\|_{L^2(P_i)}^2.$$

4.2.2. A strictly divergence-free network architecture. Instead of including a penalty term in the loss function to ensure divergence freedom in a weak sense, our third approach modifies the neural network architecture so that only solenoidal vector fields are returned. We accomplish this by learning the stream function ψ , which can be seen as a scalar potential for the divergence-free velocity vector field [14],

$$v = \nabla_{\perp} \psi = \begin{pmatrix} -\partial_y \\ \partial_x \end{pmatrix} \psi.$$

For an easy and efficient integration into the finite element framework, we are looking for divergence-free velocity corrections d^{L+1} by the neural network that are in the velocity function space, i.e., $d^{L+1} \in V_h^{L+1}$. This holds when we learn stream functions $\psi_h \in L_h^{L+1}$ that are in the pressure space L_h^{L+1} and the construction is then also again local on each patch.

For simplicity we assume that a patch P_i corresponds exactly to one element on the fine mesh Ω_{L+1} and that it is a quadrilateral parallel to the axes on an integer lattice. The nine local scalar basis functions of the second order finite elements can then be defined in the simple monomial basis,

$$L_h^{L+1}|_P = \{\xi_{i=3} \alpha + \beta \equiv \xi_{\alpha, \beta}(x, y) = x^{\alpha} y^{\beta}, \alpha, \beta = 0, 1, 2\},$$

and a stream function $\psi_h \in L_h^{L+1}$ is described by the eight basis function coefficients $s_{P_i} = (s_2, \dots, s_9)$ for this basis; the stream function has no harmonic part and $\xi_{0,0}$ can thus be disregarded. Since we work with $L_h = W_h^{(2)}$, the velocity v_h^{ψ} associated with ψ_h lies in $L_h|_P$, as desired, and it can be represented in the eight velocity basis functions derived from

$$\xi_{\alpha,\beta}(x, y) = x^\alpha y^\beta,$$

$$\begin{aligned} v_2^\psi &:= \nabla_\perp x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, & v_3^\psi &:= \nabla_\perp x^2 = \begin{pmatrix} 0 \\ 2x \end{pmatrix}, & v_4^\psi &:= \nabla_\perp y = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \\ v_5^\psi &:= \nabla_\perp xy = \begin{pmatrix} -x \\ y \end{pmatrix}, & v_6^\psi &:= \nabla_\perp x^2 y = \begin{pmatrix} -x^2 \\ 2xy \end{pmatrix}, & v_7^\psi &:= \nabla_\perp y^2 = \begin{pmatrix} -2y \\ 0 \end{pmatrix} \\ v_8^\psi &:= \nabla_\perp xy^2 = \begin{pmatrix} -2xy \\ y^2 \end{pmatrix}, & v_9^\psi &:= \nabla_\perp x^2 y^2 = \begin{pmatrix} -2x^2 y \\ 2xy^2 \end{pmatrix} \end{aligned}$$

By linearity we thus have

$$(4.4) \quad v_h^\psi|_{P_i} = \nabla_\perp \psi_h = \sum_{i=2}^9 s_i v_i^\psi$$

and the basis function coefficients for the stream function and the derived velocity are identical.

To be able to work within the usual finite element theory, we relate the non-standard velocity basis functions v_i^ψ to pointwise values using the Lagrange basis

$$(4.5) \quad V_h|_P = \{\phi_1^x, \dots, \phi_9^x, \phi_1^y, \dots, \phi_9^y\}$$

that satisfies

$$\phi_i^x(x_j) = \begin{pmatrix} \delta_{ij} \\ 0 \end{pmatrix}, \quad \phi_i^y(x_j) = \begin{pmatrix} 0 \\ \delta_{ij} \end{pmatrix}, \quad i, j = 1, \dots, 9$$

for the nine Lagrange points δ_{ij} . Each of the derived velocity functions v_i^ψ has a unique representation in this basis with coefficients η_{ij}^x, η_{ij}^y for $i = 2, \dots, 9$ and $j = 1, \dots, 9$. The pointwise values of the velocity vector field can thus be reconstructed as

$$(4.6) \quad v_h^\psi|_{P_i} = \sum_{k=2}^9 s_k \sum_{l=1}^9 \eta_{kl}^x \phi_l^x + \eta_{kl}^y \phi_l^y.$$

Along boundaries of the domain $\partial\Omega$ where Dirichlet conditions are prescribed, the corresponding coefficients η_{kl}^x, η_{kl}^y are set to zero such that the resulting update v_h^ψ is no longer strictly divergence free.

With the formulation above, we can train for the eight stream function coefficients $s_{P_i} = (s_2, \dots, s_9)$ that through (4.4) and (4.6) correspond to a divergence-free velocity correction $d^{L+1}(\{s_i^{L+1}\})$ satisfying $\|\nabla \cdot d^{L+1}\|_{L^2(\Omega)} = 0$. Similar to the DNN-MG setting, the corrections $d^{L+1}(\{s_i^{L+1}\})$ are averaged over all degrees of freedom shared by multiple patches. For network training, we minimize

$$\mathcal{L}_\psi(v^L, v^{L+1}; s^{L+1}) := \sum_{P_j \in \Omega_h} \|v^{L+1} - (v^L + d^{L+1}(\{s_i^{L+1}\}))\|_{L^2(P_j)}^2.$$

4.3. Numerical evaluation. To evaluate the efficacy of the three approaches developed in Section 4.2 to improve the divergence freedom of the velocity vector field, we implemented them using PyTorch [41] and Gascoigne 3D [4] based on the framework developed in [36]. For the penalty-based approaches with the loss functions \mathcal{L}_{p_1} and \mathcal{L}_{p_2} the network and other aspects of the training process remained unchanged compared to [36], where the network had 8634 parameters. The network architecture we use for the stream function-based approach

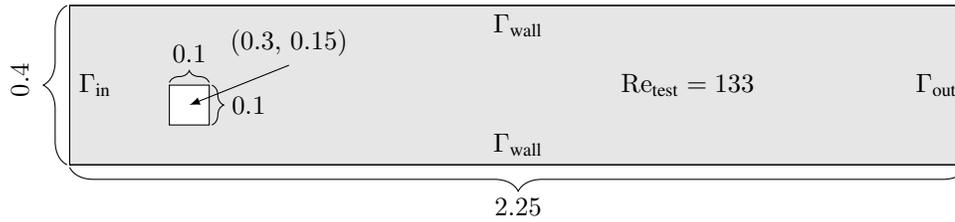


FIG. 4.2. Geometry of the training scenario with a parabolic inflow profile Γ_{in} , do-nothing boundary conditions at the outflow boundary Γ_{out} and no-slip conditions on the walls Γ_{wall} . The center of the obstacle is at $(0.3, 0.15)$. For the test scenario the obstacle is shifted such that the center is at $(0.3, 0.25)$.

with \mathcal{L}_ψ is also similar to the one in [36] but the convolutional layers are replaced by fully connected ones of size 8×8 . This results in a network with 9000 parameters.

The flow we consider in our experiments is a modification of the classical benchmark of a flow around a cylinder [50] where the obstacle is now a square; see Figure 4.2. The modification simplifies the implementation, since we are able to use uniform meshes, but does not change the qualitative behavior of the quantities of interest to us and in particular one still has a non-stationary wake behind the obstacle as in the original benchmark [50]. The singularities at the corners of the obstacles have been excluded in the evaluation below.

Divergence freedom. We first compare the divergence $\|\nabla \cdot v(t)\|_{L_2(\Omega_L)}$ obtained with the three approaches (DNN-MG- \mathcal{L}_i) as well as a coarse multigrid solution (MG), the original DNN-MG method proposed in [36] (DNN-MG), and, as reference, a fine multigrid solution on level $L + 1$ (MG-fine). The results are presented in Figure 4.1. They show that the approaches from Section 4.2 are not able to significantly reduce the divergence. While DNN-MG- \mathcal{L}_ψ based on the strictly divergence-free output performs best, the reduction is still modest. Ultimately, however, this is not surprising since

$$\|\operatorname{div} d^{L+1}\| = 0 \quad \Rightarrow \quad \|\operatorname{div} (v^L + d^{L+1})\| = \|\operatorname{div} v^L\|,$$

and the perfectly divergence-free correction does not address the divergence residual in v^L , which is substantial. The second approach, using \mathcal{L}_{p_2} , is designed to address this but no reduction can be observed in practice.

We investigated the disappointing performance of the modified loss functions \mathcal{L}_{p_1} and \mathcal{L}_{p_2} that include a penalty term and found that it stems from an unstable balance between the minimization of the divergence and the fulfillment of the finite element Galerkin equation: If the weighting factor γ for the divergence is chosen close to zero then the network behaves like the classical DNN-MG approach. If we choose larger values for $\gamma > 0$, however, then the network is not able to sufficiently minimize the discrepancy to the high fidelity Galerkin solution and overall the approximation quality suffers. We tested γ in the range $10^{-1}, \dots, 10^{-10}$ but did not find a value where the two error terms in the loss functions \mathcal{L}_{p_1} and \mathcal{L}_{p_2} are well balanced. For $\gamma > 10^{-7}$ we, in fact, observed that the penalty term causes the output to converge towards zero, which is a divergence-free velocity vector field but not the desired correction. The effect is most pronounced for DNN-MG- \mathcal{L}_{p_2} .

Drag and Lift functionals. Figure 4.3 displays the drag and lift functionals of the obstacle for the different approaches. In Table 4.1 we also show the corresponding minimum and maximum values, the mean values, and the amplitudes of the oscillation as well as the frequencies of the periodic flow pattern. In the lower table of Table 4.1 we report the relative errors with respect to the high fidelity solution MG (fine).

The results show that the non-smooth corners of the square obstacle have a negative effect on the performance of the original DNN-MG approach since it no longer provides the

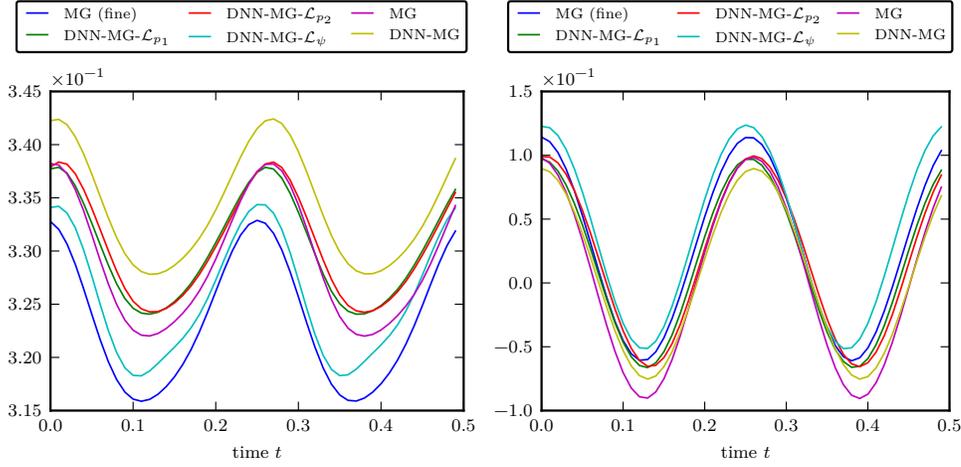


FIG. 4.3. Comparison of drag and lift functionals for the coarse mesh (MG), the high fidelity (MG fine) and the deep neural network multigrid solution (DNN-MG). Three results of the attempts to enforce divergence freedom are shown: DNN-MG with the divergence of the DNN’s output as penalty term (DNN-MG- \mathcal{L}_{p_1}), DNN-MG with the divergence of the corrected velocity field as penalty term (DNN-MG- \mathcal{L}_{p_2}) and enforcing divergence freedom by construction (DNN-MG- \mathcal{L}_{ψ}).

improvement in the functional values observed in [36] for a smooth obstacle. Among the approaches proposed in the present work, DNN-MG- \mathcal{L}_{ψ} based on learning the stream function provides the most promising results. It is able to reduce the relative error in all functional outputs and almost matches the values of MG (fine) in Figure 4.3. Furthermore, the training time for the neural network is in this case only $\approx 75\%$ in comparison to the other DNN-MG approaches.

For DNN-MG- \mathcal{L}_{ψ} we observe a deviation of the oscillation frequency that is larger than for the other approaches. We believe that this defect stems from a complicated interplay of temporal dissipation and spatial discretization. For a discussion we refer to [36] and also to [37] where the same effect is observed for a parallel time stepping scheme. In the context of the deep neural network multigrid approach it must be investigated if this problem can be resolved by training with high fidelity solutions that come from discretizations that are finer in space and in time, i.e., by working with a finer time step.

Velocity vector field. Figure 4.4 visualizes the error $|v(x) - v_{\text{fine}}(x)|$ in the velocity vector field for MG, DNN-MG, DNN-MG- \mathcal{L}_{p_1} , DNN-MG- \mathcal{L}_{p_2} , and DNN-MG- \mathcal{L}_{ψ} with respect to the fine reference solution MG (fine). In Figure 4.5 we see a comparison of velocity errors measured in the Euclidean norm of the discrete solution. To obtain meaningful comparisons, the outputs are synchronized such that we consider snapshots starting from a maximum of the drag functional; cf. Figure 4.3. The results in Figure 4.5 verify that all approaches proposed in Section 4.2 are able to reduce the error in the velocity field compared to MG and DNN-MG. The results differ, however, from those for the divergence freedom in Figure 4.1 in that DNN-MG- \mathcal{L}_{p_1} and DNN-MG- \mathcal{L}_{p_2} provide better results for the velocity vector field than the stream function formulation DNN-MG- \mathcal{L}_{ψ} . We attribute this to the larger prediction space of DNN-MG- \mathcal{L}_{p_1} and DNN-MG- \mathcal{L}_{p_2} compared to the reduced divergence-free space that is used in DNN-MG- \mathcal{L}_{ψ} .

Discussion. The results presented in this section are surprising in that the loss functions of Section 4.2 designed to reduce the divergence improve the overall quality of the obtained results, e.g., in terms of the lift and drag functionals, but do not lead to significantly better

TABLE 4.1

Maximum, minimum values, mean and amplitude of oscillation for drag and lift functionals as well as frequency of the periodic solution. We indicate the results for the different finite element solution on coarse and fine mesh finite element simulation and for the different hybrid finite element / deep neural network approaches. While the upper table shows the raw results, the lower table indicates the relative error (in %) with respect to the high fidelity solution. For each value we highlight the best result.

Approach	Drag				Lift				Freq.
	Min	Max	Mean	Ampl.	Min	Max	Mean	Ampl.	
coarse	0.3220	0.3382	0.3301	0.0162	-0.0905	0.0985	0.0040	0.1890	3.8095
fine	0.3159	0.3329	0.3244	0.0170	-0.0610	0.1144	0.0267	0.1754	3.9024
DNN-MG	0.3277	0.3425	0.3351	0.0148	-0.0753	0.0896	0.0071	0.1650	3.8298
DNN-MG- \mathcal{L}_{p_1}	0.3240	0.3379	0.3309	0.0139	-0.0665	0.0978	0.0156	0.1643	3.9130
DNN-MG- \mathcal{L}_{p_2}	0.3242	0.3384	0.3313	0.0142	-0.0657	0.0999	0.0171	0.1657	3.9130
DNN-MG- \mathcal{L}_ψ	0.3178	0.3350	0.3264	0.0172	-0.0544	0.1259	0.0357	0.1803	4.0179

Approach	Drag				Lift				Freq.
	Min	Max	Mean	Ampl.	Min	Max	Mean	Ampl.	
coarse	1.93	1.59	1.76	4.71	48.4	13.9	85.0	7.75	2.38
DNN-MG	3.74	2.88	3.30	12.9	23.4	21.7	73.4	5.93	1.86
DNN-MG- \mathcal{L}_{p_1}	2.56	1.50	2.00	18.2	9.02	14.5	41.6	6.33	0.27
DNN-MG- \mathcal{L}_{p_2}	2.63	1.65	2.13	16.5	7.70	12.7	36.0	5.53	0.27
DNN-MG- \mathcal{L}_ψ	0.60	0.63	0.62	1.18	10.8	10.1	33.7	2.79	2.96

divergence freedom. More work is hence required to obtain hybrid finite element/neural network-based simulations that improve this important property. For the loss functions with the penalty factor it should in particular be reconsidered how these can be formulated so that their two constitutive terms do not compete with each other as we observed in our experiments.

The efficiency of DNN-MG- \mathcal{L}_ψ based on the stream function is not entirely surprising to us. The network there only has to learn a scalar function instead of two components of a vector field that, furthermore, satisfy a nontrivial constraint through the divergence freedom. The stream function formulation also encodes the structure of the dynamics in a different way than the usual velocity one, cf. [39], and this leads often to simulations that preserve structural properties such as energy or vorticity. In the finite element context this was already observed by Fix [19]. We believe that this partly also explains the reduced training time for DNN-MG- \mathcal{L}_ψ .

5. Conclusion. We have presented three approaches to enforce divergence freedom within the hybrid finite element / neural network DNN-MG method [36] for the simulation of the Navier-Stokes equations. The first two of these use a penalty term that encourages divergence freedom while the third one guarantees a divergence-free correction by learning the stream function. The penalty term-based approaches did not provide significant improvements compared to DNN-MG for both the divergence and the drag and lift functionals. The third approach based on the stream function yielded significantly better overall results but also only a rather modest improvement for the divergence.

More work is hence required in the future to obtain hybrid finite element / neural network simulations that respect the divergence freedom of the Navier-Stokes equations. For the penalty-based approaches it should be investigated how the two terms in the loss function can be balanced. DNN-MG- \mathcal{L}_2 is thereby particularly interesting since it allows, in principle, to provide an overall reduction by correcting v^L towards a lower divergence. For DNN-MG-

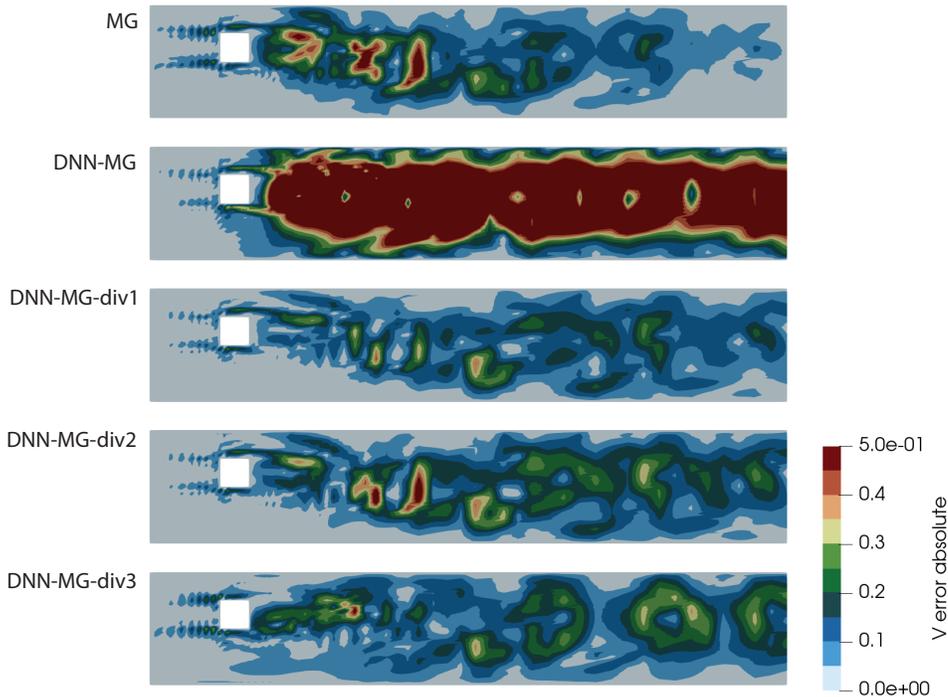


FIG. 4.4. Velocity error for the different approaches with respect to the high fidelity solution MG (fine).

\mathcal{L}_ψ it would be interesting to investigate the efficiency when the velocity-stream function formulation is also used for the simulation.

In future work, we also want to investigate a hybrid finite element / neural network simulation for the Navier-Stokes equations in 3D. This is, however, considerably more challenging. For example, the stream function / vorticity formulation, which proved particular effective in 2D, does not exist in the same simple form there. One alternative would be to use a structure preserving discretization of the Navier-Stokes equations, e.g., [20, 21], as a basis. The built-in constraints, such as an inherently divergence-free velocity vector field, would then again aid the physicality of the learned solution and simplify the learning task. How other physical invariants, e.g., energy or Kelvin’s circulation theorem, can be incorporated into hybrid simulations and what benefits this provides should also be investigated in more detail. Stability, consistency, and convergence are, furthermore, pertinent open questions that should be addressed.

Acknowledgments. NM and TR acknowledge the financial support by the Federal Ministry of Education and Research of Germany, grant number 05M16NMA as well as the GRK 2297 MathCoRe, funded by the Deutsche Forschungsgemeinschaft, grant number 314838170. NM acknowledges support by the Helmholtz-Gesellschaft grant number HIDSS-0002 DASHH. TR further acknowledges funding by the Deutsche Forschungsgemeinschaft, grant number 411046898. CL is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 422037413 – TRR 287.

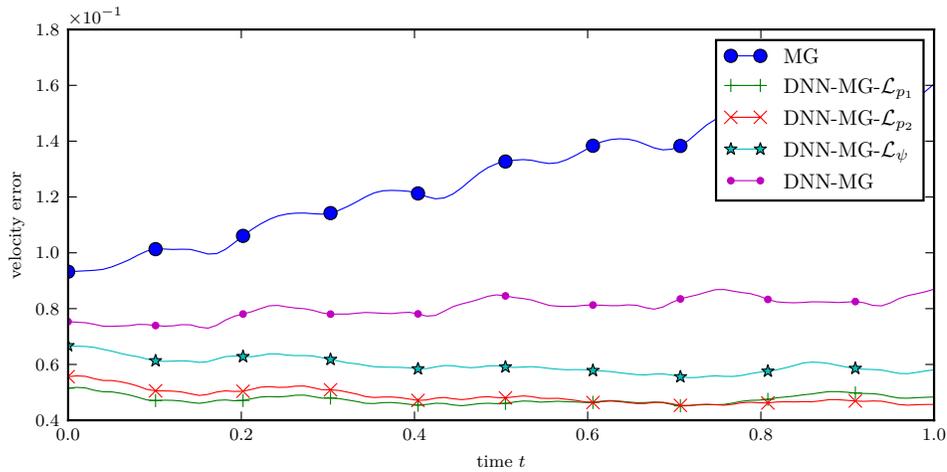


FIG. 4.5. Relative velocity error in the l^2 -norm for the different approaches with respect to the high fidelity solution MG (fine).

REFERENCES

- [1] S. BAI, J. Z. KOLTER, AND V. KOLTUN, *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*, Preprint, arXiv:1803.01271, March 2018. <https://arxiv.org/abs/1803.01271>.
- [2] Y. BAR-SINAI, S. HOYER, J. HICKEY, AND M. P. BRENNER, *Learning data-driven discretizations for partial differential equations*, Proc. Nat. Acad. Sci. USA, 116 (2019), pp. 15344–15349.
- [3] R. BECKER AND M. BRAACK, *A finite element pressure gradient stabilization for the Stokes equations based on local projections*, Calcolo, 38 (2001), pp. 173–199.
- [4] R. BECKER, M. BRAACK, D. MEIDNER, T. RICHTER, AND B. VEXLER, *The finite element toolkit GASCOIGNE*. <http://www.gascoigne.de>.
- [5] E. J. BEKKERS, *B-spline CNNs on Lie groups*, in International Conference on Learning Representations, OpenReview.net, 2020.
- [6] J. BERG AND K. NYSTRÖM, *Data-driven discovery of PDEs in complex datasets*, J. Comput. Phys., 384 (2019), pp. 239–252.
- [7] K. BHATTACHARYA, B. HOSSEINI, N. B. KOVACHKI, AND A. M. STUART, *Model reduction and neural networks for parametric PDEs*, J. Comput. Math., 7 (2021), pp. 121–157.
- [8] A. BIETTI AND J. MAIRAL, *Group invariance, stability to deformations, and complexity of deep convolutional representations*, J. Mach. Learn. Res., 20 (2019), pp. 876–924.
- [9] M. BRAACK AND P.B. MUCHA, *Directional do-nothing condition for the Navier-Stokes equations*, J. Comp. Math., 32 (2014), pp. 507–521.
- [10] J. BRUNA AND S. MALLAT, *Invariant scattering convolution networks*, IEEE Trans. Patt. Anal. Machine Intell., 35 (2013), pp. 1872–1886.
- [11] T. CHEN AND H. CHEN, *Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems*, IEEE Trans. Neural Netw., 6 (1995), pp. 911–917.
- [12] Z. CHEN, J. ZHANG, M. ARJOVSKY, AND L. BOTTOU, *Symplectic recurrent neural networks*, Preprint, arXiv:1909.13334, September 2019. <https://arxiv.org/abs/1909.13334>.
- [13] K. CHO, B. VAN MERRIENBOER, C. GULCEHRE, F. BOUGARES, H. SCHWENK, AND Y. BENGIO, *Learning phrase representations using RNN encoder-decoder for statistical machine translation*, in Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), ACL, Stroudsburg, 2014, pp. 1724–1734.
- [14] A. J. CHORIN AND J. E. MARSDEN, *A Mathematical Introduction to Fluid Mechanics*, 3rd. ed., Springer, New York, 1993.
- [15] M. CRANMER, S. GREYDANUS, S. HOYER, P. BATTAGLIA, D. SPERGEL, AND S. HO, *Lagrangian neural networks*, Preprint, arXiv:2003.04630, March 2020. <https://arxiv.org/abs/2003.04630>.
- [16] W. E AND B. YU, *The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems*, Commun. Math. Stat., 6 (2018), pp. 1–12.

- [17] L. FAILER AND T. RICHTER, *A Newton multigrid framework for optimal control of fluid-structure interactions*, *Optim. Eng.*, 22 (2021), pp. 2009–2037.
- [18] M. FINZI, K. A. WANG, AND A. G. WILSON, *Simplifying Hamiltonian and Lagrangian neural networks via explicit constraints*, in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, H. Larochell, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, eds., 2020.
- [19] G. J. FIX, *Finite element models for ocean circulation problems*, *SIAM J. Appl. Math.*, 29 (1975), pp. 371–387.
- [20] E. S. GAWLIK AND F. GAY-BALMAZ, *A conservative finite element method for the incompressible Euler equations with variable density*, *J. Comput. Phys.*, 412 (2020), Art. 109439, 22 pages.
- [21] ———, *A variational finite element discretization of compressible flow*, *Found. Comput. Math.*, 21 (2021), pp. 961–1001.
- [22] J. HAN, A. JENTZEN, AND W. E., *Solving high-dimensional partial differential equations using deep learning*, *Proc. Nat. Acad. Sci. USA*, 115 (2018), pp. 8505–8510.
- [23] J.G. HEYWOOD, R. RANNACHER, AND S. TUREK, *Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations*, *Internat. J. Numer. Meth. Fluids.*, 22 (1996), pp. 325–352.
- [24] S. HOCHREITER AND J. SCHMIDHUBER, *Long Short-Term Memory*, *Neural Comput.*, 9 (1997), pp. 1735–1780.
- [25] Y. HU, T. ZHAO, Z. XU, AND L. LIN, *Neural-PDE: a RNN based neural network for solving time dependent PDEs*, Preprint, arXiv:2009.03892, September 2020. <https://arxiv.org/abs/2009.03892>.
- [26] P. JIN, A. ZHU, G. E. KARNIADAKIS, AND Y. TANG, *SympNets: intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems*, Preprint, arXiv:2001.03750, January 2020. <https://arxiv.org/abs/2001.03750>.
- [27] M. F. KASIM, D. WATSON-PARRIS, L. DEACONU, S. OLIVER, P. HATFIELD, D. H. FROULA, G. GREGORI, M. JARVIS, S. KHATIWALA, J. KORENAGA, J. TOPP-MUGGLESTONE, E. VIEZZER, AND S. M. VINKO, *Building high accuracy emulators for scientific simulations with deep neural architecture search*, Preprint, arXiv:2001.08055, January 2020. <https://arxiv.org/abs/2001.08055>.
- [28] Y. LECUN, Y. BENGIO, AND G. HINTON, *Deep learning*, *Nature*, 521 (2015), pp. 436–444.
- [29] Z. LI, N. KOVACHKI, K. AZIZZADENESHELI, B. LIU, K. BHATTACHARYA, A. STUART, AND A. ANANDKUMAR, *Fourier neural operator for parametric partial differential equations*, Preprint, arXiv:2010.08895, October 2020. <https://arxiv.org/abs/2010.08895>.
- [30] Z. LI, N. KOVACHKI, K. AZIZZADENESHELI, B. LIU, K. BHATTACHARYA, A. STUART, AND A. ANANDKUMAR, *Neural operator: graph kernel network for partial differential equations*, Preprint, arXiv:2003.03485, March 2020. <https://arxiv.org/abs/2003.03485>.
- [31] J. LING, A. KURZAWSKI, AND J. TEMPLETON, *Reynolds averaged turbulence modelling using deep neural networks with embedded invariance*, *J. Fluid Mech.*, 807 (2016), pp. 155–166.
- [32] Z. LONG, Y. LU, AND B. DONG, *PDE-net 2.0: learning PDEs from data with a numeric-symbolic hybrid deep network*, *J. Comput. Phys.*, 399 (2019), Art. 108925, 18 pages.
- [33] L. LU, P. JIN, AND G. E. KARNIADAKIS, *DeepONet: learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators*, Preprint, arXiv:1910.03193, October 2019. <https://arxiv.org/abs/1910.03193>.
- [34] L. LU, X. MENG, Z. MAO, AND G. E. KARNIADAKIS, *DeepXDE: a deep learning library for solving differential equations*, Preprint, arXiv:1907.04502, July 2019. <https://arxiv.org/abs/1907.04502>.
- [35] A. KLAWONN, M. EICHINGER, AND A. HEINLEIN, *Stationary flow predictions using convolutional neural networks*, in F. Vermolen and C. Vuik, eds., *Numerical Mathematics and Advanced Applications ENUMATH 2019*, Springer, Basel, 2020.
- [36] N. MARGENBERG, D. HARTMANN, C. LESSIG, AND T. RICHTER, *A neural network multigrid solver for the Navier-Stokes equations*, Preprint, arXiv:2008.11520, August 2020. <https://arxiv.org/abs/2008.11520>.
- [37] N. MARGENBERG AND T. RICHTER, *Parallel time-stepping for fluid-structure interactions*, *Math. Model. Nat. Phenom.*, 16 (2021), Art. 20, 19 pages.
- [38] M. MATTHEAKIS, P. PROTOPAPAS, D. SONDAK, M. DI GIOVANNI, AND E. KAXIRAS, *Physical symmetries embedded in neural networks*, Preprint, arXiv:1904.08991, April 2019. <https://arxiv.org/abs/1904.08991>.
- [39] P. J. MORRISON, *Hamiltonian description of the ideal fluid*, *Rev. Modern Phys.*, 70 (1998), pp. 467–521.
- [40] M. A. NABIAN AND H. MEIDANI, *Physics-driven regularization of deep neural networks for enhanced engineering design and analysis*, Preprint, arXiv:1810.05547, October 2018. <https://arxiv.org/abs/1810.05547>.
- [41] A. PASZKE, S. GROSS, F. MASSA, A. LERER, J. BRADBURY, G. CHANAN, T. KILLEEN, Z. LIN, N. GIMELSHEIN, L. ANTIGA, A. DESMAISON, A. KOPF, E. YANG, Z. DEVITO, M. RAISON, A. TEJANI, S. CHILAMKURTHY, B. STEINER, L. FANG, J. BAI, AND S. CHINTALA, *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, and R. Garnett, ed., Curran Associates,

- Red Hook, 2019, pp. 8024–8035.
- [42] M. RAISSI AND G. E. KARNIADAKIS, *Hidden physics models: machine learning of nonlinear partial differential equations*, J. Comput. Phys., 357 (2018), pp. 125–141.
 - [43] M. RAISSI, P. PERDIKARIS, AND G.E. KARNIADAKIS, *Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys., 378 (2019), pp. 686–707.
 - [44] M. RAISSI, A. YAZDANI, AND G. E. KARNIADAKIS, *Hidden fluid mechanics: learning velocity and pressure fields from flow visualizations*, Science, 367 (2020), pp. 1026–1030.
 - [45] T. RICHTER, *Fluid-Structure Interactions. Models, Analysis and Finite Elements*, Springer, Basel, 2017.
 - [46] S. RUDY, A. ALLA, S. L. BRUNTON, AND J. N. KUTZ, *Data-driven identification of parametric partial differential equations*, SIAM J. Appl. Dyn. Syst., 18 (2019), pp. 643–660.
 - [47] L. RUTHOTTO AND E. HABER, *Deep neural networks motivated by partial differential equations*, J. Math. Imaging Vision, 62 (2020), pp. 352–364.
 - [48] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, PWS Publishing Company, Boston, 1996.
 - [49] T. SAPSIS, Z. Y. WAN, B. DODOV, H. A. DIJKSTRA, AND C. LESSIG, *Data-assisted reduced-order modeling of climate dynamics*, in Proceedings of EGU2019, orals, 2019.
 - [50] M. SCHÄFER, S. TUREK, F. DURST, E. KRAUSE, AND R. RANNACHER, *Benchmark computations of laminar flow around a cylinder*, in Flow Simulation with High-Performance Computers II, E. H. Hirschel, ed., Notes Numer. Fluid Mech., vol. 48, Vieweg, Wiesbaden, 1996, pp 547–566.
 - [51] B. STEVENS AND T. COLONIUS, *Finitenet: a fully convolutional LSTM network architecture for time-dependent partial differential equations*, in 73rd Annual Meeting of the APS Division of Fluid Dynamics, Bull Am Phys. Soc, vol. 65, Abstract: R01.00014, 2020.
 - [52] A. M. TARTAKOVSKY, C. O. MARRERO, P. PERDIKARIS, G. D. TARTAKOVSKY, AND D. BARAJAS-SOLANO, *Learning parameters and constitutive relationships with physics informed deep neural networks*, Preprint, arXiv:1808.03398, August 2018. <https://arxiv.org/abs/1808.03398>.
 - [53] S. P. VANKA, *Block-implicit multigrid solution of Navier-Stokes equations in primitive variables*, J. Comp. Phys., 65 (1985), pp. 138–158.
 - [54] Z. Y. WAN, B. DODOV, C. LESSIG, H. DIJKSTRA, AND T. P. SAPSIS, *A data-driven framework for the stochastic reconstruction of small-scale features in climate data sets*, J. Comput. Phys, 442, (2021), Art. 110484, 24 pages.
 - [55] S. WIEWEL, M. BECHER, AND N. THUEREY, *Latent-space physics: towards learning the temporal evolution of fluid flow*, Preprint, arXiv:1802.10123, February 2018. <https://arxiv.org/abs/1802.10123>.
 - [56] Y. YANG AND P. PERDIKARIS, *Physics-informed deep generative models*, Preprint, arXiv:1812.03511, December 2018. <https://arxiv.org/abs/1812.03511>.