

# RESEARCH ANNOUNCEMENTS ON “DPK: DEEP NEURAL NETWORK APPROXIMATION OF THE FIRST PIOLA-KIRCHHOFF STRESS”

YANG Zhi-jian, YUAN Cheng

(School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China)

## 1 Introduction and Main Results

Consider the first Piola-Kirchhoff stress  $\mathbf{P}$  defined as [1]

$$\mathbf{P}(\mathbf{F}) = \frac{\partial W(\mathbf{F})}{\partial \mathbf{F}}, \quad (1.1)$$

where  $\mathbf{F}$  and  $W$  refer to the deformation tensor and strain energy density. Due to the principle of material frame-indifference and the material symmetry, two important constraints should be satisfied by these relations [1]:

$$\mathbf{P}(\mathbf{QF}) = \mathbf{Q}\mathbf{P}(\mathbf{F}), \quad \forall \mathbf{Q} \in SO(3); \quad (1.2)$$

$$\mathbf{P}(\mathbf{F}) = \mathbf{P}(\mathbf{F}\mathbf{H})\mathbf{H}^T, \quad \forall \mathbf{H} \in G, \quad (1.3)$$

where  $SO(3)$  stands for the proper orthogonal group and  $G$  refers to the material symmetry group.

In this work, we propose a deep symmetric network  $\mathbf{P}_\theta(\mathbf{F})$  to represent the first PK stress (DPK). By symmetric we mean the network can exactly satisfy the two constraints (1.2),(1.3). The neural network enables us to construct the constitutive relation based on both macroscopic observations and atomistic simulation data.

To begin with, we first review the traditional atomistic PK stress (virial stress) defined as

$$\mathbf{P}(\mathbf{F}) = \frac{1}{\Omega} \frac{\partial V(|\mathbf{F}\mathbf{X}_{12}|, \dots, |\mathbf{F}\mathbf{X}_{ij}|, \dots)}{\partial \mathbf{F}} = -\frac{1}{2\Omega} \sum_i \sum_{j \neq i} \mathbf{f}_{ij} \otimes \mathbf{X}_{ij}, \quad (1.4)$$

in which  $\mathbf{X}_{ij}$  and  $\mathbf{f}_{ij}$  refer to the relative displacement and pair force between the  $i$ -th and  $j$ -th atom, and  $V$  stands for the potential function of atom system. Under several assumption on the symmetry of the potential and the crystal structure, we can prove the virial stress is exactly in conformity with the two constraints (1.2),(1.3):

---

**Foundation item:** Supported by National Key Research and Development Program of China (2020YFA0714200), National Nature Science Foundation of China (12125103; 12071362), Natural Science Foundation of Hubei Province (2021AAA010; 2019CFA007) and Fundamental Research Funds for the Central Universities.

**Theorem 1.1 (Symmetry of virial stress)** For a given crystal system with periodic boundary condition, suppose the potential  $V$  is a symmetric function, let  $\mathbf{P}(\mathbf{F})$  be the stress defined by (1.4), then the constraint (1.2) holds for any  $\mathbf{Q}$  in  $SO(3)$ , and (1.3) holds for any  $\mathbf{H}$  in the crystallographic symmetry group (point symmetry group  $G_p$ , [2]).

Motivated by the above theorem, we design the deep symmetry network as follows. For a given crystal material, we denote the crystallographic point symmetry group as  $G_p$ . For any finite 3-dimensional trainable weight vectors  $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ , the symmetrized weights  $\tilde{\mathbf{W}}$  is defined as the image set of  $G_p$  on  $\mathbf{W}$ :

$$\tilde{\mathbf{W}} = \{\mathbf{H}_i \mathbf{w}_j | \mathbf{H}_i \in G_p, \mathbf{w}_j \in \mathbf{W}\}. \quad (1.5)$$

By construction  $\tilde{\mathbf{W}} = \{\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_M\}$  is an invariant subset of  $G_p$ . With this we define the deep symmetry network as follows:

**Definition 1.1 (DPK)** Let  $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$  be  $N$  trainable 3-dimensional weight vectors and  $\tilde{\mathbf{W}}$  the symmetrized weights defined above, denote  $g(\mathbf{x}; \theta) = h_n \circ a_{n-1} \circ \dots \circ a_1 \circ h_1(\mathbf{x})$  as the classical dense neural network (DNN) with linear functions  $h_i(\mathbf{x}; \theta_i)$  and activation function  $a_i(\mathbf{x})$ , the deep symmetric network for PK stress is defined as

$$\mathbf{P}_{\theta, \mathbf{W}}(\mathbf{F}) = \mathbf{F} \sum_{i=1}^M \frac{\partial f}{\partial x_i}(|\mathbf{F}\tilde{\mathbf{w}}_1|, \dots, |\mathbf{F}\tilde{\mathbf{w}}_M|; \theta) \frac{1}{|\mathbf{F}\tilde{\mathbf{w}}_i|} \tilde{\mathbf{w}}_i \otimes \tilde{\mathbf{w}}_i, \quad (1.6)$$

where  $f(\mathbf{x}, \theta) = \frac{1}{M!} \sum_{\tau \in S_M} g(\tau(\mathbf{x}); \theta)$  with  $S_M$  stands for all the permutations of  $[1, 2, \dots, M]$ .

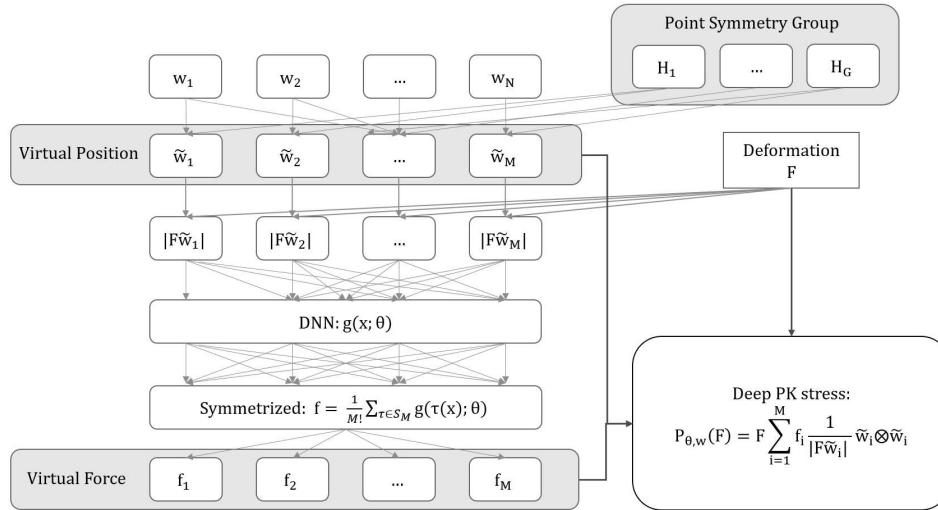


Figure 1 The network architecture of DPK.

The architecture of this network is illustrated in Fig.1. By using the fact that  $\tilde{\mathbf{W}}$  is an invariant subset of  $G_p$  and  $f(\mathbf{x}, \theta) = \frac{1}{M!} \sum_{\tau \in S_M} g(\tau(\mathbf{x}); \theta)$  is a symmetric function, we obtain that the DPK is a symmetric stress:

**Theorem 1.2 (Symmetry of DPK)** For a given crystal system with crystallographic symmetry group  $G_p$ , let  $\mathbf{P}_{\theta, \mathbf{w}}(\mathbf{F})$  be the DPK stress defined by (1.6), then the constraint (1.2) holds for any  $\mathbf{Q}$  in  $SO(3)$ , and (1.3) holds for any  $\mathbf{H}$  in  $G_p$ .

Furthermore, by applying some existing result on the approximation of DNN, we show that the DPK stress owns good approximation ability:

**Theorem 1.3 (Approximation of DPK)** For a finite crystal system consists of  $N$  atoms with symmetric potential  $V$  satisfying  $\|\mathbf{X}_{ij}\|_2 < 1$  for any  $j \neq i$ , if there exists some  $n > 1, B < \infty$  such that  $\|V\|_{W^{n, \infty}((0,1)^{N(N-1)})} < B$ , then for any  $\epsilon \in (0, 1/2)$  there exists a DPK function  $\mathbf{P}_0$  such that

$$\|\mathbf{P}_{virial}(\mathbf{F}) - \mathbf{P}_0(\mathbf{F})\|_F \leq \epsilon, \quad (1.7)$$

if  $\|\mathbf{F}\mathbf{X}_{ij}\|_2 < 1$ . Here  $\|\cdot\|_F$  represents the Frobenius norm and  $P_{virial}(\mathbf{F})$  is the virial stress defined by (1.4).

**Remark 1** Although the DPK architecture introduced above supplies us with a strong approximation tool which can exactly satisfy the symmetry conditions, the sum term  $f(\mathbf{x}, \theta) = \frac{1}{M!} \sum_{\tau \in S_M} g(\tau(\mathbf{x}); \theta)$  with a large  $M$  will introduce an intolerable computation complexity in both training and inference procedure.

In order to achieve a simplified architecture while keeping its symmetry and approximation capability, we learn from the idea of quasi-harmonic approximation [3] method and decompose the virtual potential function  $f$  in (1.6) as

$$f(x_1, x_2, \dots, x_M; \theta) = \sum_{i=1}^M g(x_i; \theta) \quad (1.8)$$

where  $g(x) = h_n \circ a_{n-1} \circ \dots \circ a_1 \circ h_1(x)$  is the classical DNN function. Applying (1.8) to (1.6), we can obtain the modified DPK architecture as:

$$\mathbf{P}_{\theta, \mathbf{w}}(\mathbf{F}) = \mathbf{F} \sum_{i=1}^M \frac{dg}{dx}(|\mathbf{F}\tilde{\mathbf{w}}_i|; \theta) \frac{1}{|\mathbf{F}\tilde{\mathbf{w}}_i|} \tilde{\mathbf{w}}_i \otimes \tilde{\mathbf{w}}_i. \quad (1.9)$$

A further simplification can be achieved once we represent  $\frac{1}{x}g'(x; \theta) = \tilde{g}(x; \tilde{\theta})$  where  $\tilde{g}(x; \tilde{\theta})$  is a new DNN function. In this case the simplified DPK can be written as:

$$\mathbf{P}_{\theta, \mathbf{w}}(\mathbf{F}) = \mathbf{F} \sum_{i=1}^M \tilde{g}(|\mathbf{F}\tilde{\mathbf{w}}_i|; \tilde{\theta}) \tilde{\mathbf{w}}_i \otimes \tilde{\mathbf{w}}_i. \quad (1.10)$$

We numerically illustrate the effectiveness and transferability of (1.10). Additionally, with some classical tools in learning theory such as fat-shattering dimension, we prove that the estimation error (generalization error) of (1.10) can be arbitrarily small with sufficient large number of training samples:

**Theorem 1.4 (Learnability of simplified DPK)** Let  $\mathcal{A} = \{\mathbf{P}_{\theta, \mathbf{w}}\}$  be the class of DPK functions defined by (1.10) satisfying that  $\tilde{g}(x; \tilde{\theta})$  has  $l \geq 2$  layers,  $N_w$  weights (bounded by  $V$ ), Lipschitz continuous activation function with Lipschitz coefficient  $L > 1/V$  and each unit maps into  $[-b, b]$  for some  $b > 0$ . Given the training samples  $S_m = \{(\mathbf{F}^1, \mathbf{P}^1), \dots, (\mathbf{F}^m, \mathbf{P}^m)\}$  according to independent identical distribution  $D$  on  $[-1, 1]^{3 \times 3} \times [-1, 1]^{3 \times 3}$  and loss function

$$l(f(\mathbf{X}), \mathbf{Y}) = \frac{1}{9} \sum_{i,j} l_0(f_{ij}(\mathbf{X}), \mathbf{Y}_{ij}), \text{ for } f \in \mathcal{A}, \quad (1.11)$$

where  $l_0(x, y) = |x - y|$ , then for any approximate sample error minimization (SME, [4]) algorithm  $\mathbb{M}$  and  $\epsilon \leq 2 \max\{b, V, \sqrt{3V}\}, \delta > 0$ , there exists some  $m_0(\epsilon, \delta, \mathcal{A}) < \infty$  such that  $\forall m > m_0$ , with probability at least  $1 - \delta$ ,  $\mathbb{M}$  returns a function  $P^* \in \mathcal{A}$  satisfying

$$E_{(x,y) \sim D}[l(P^*(x), y)] < \inf_{f \in \mathcal{A}} E_{(x,y) \sim D}[l(f(x), y)] + \epsilon. \quad (1.12)$$

**Remark 2** In practice, the conditions of  $\tilde{g}$  are indeed satisfied since we will use the ReLU as activation function and bound the network weights by using an  $l_1$  regularization.

The detailed proof of the theorems mentioned above and the numerical results can be found in [5].

## References

- [1] Tadmor B Ellad, Miller E Ronald. Modeling materials: continuum, atomistic and multiscale techniques[M]. Cambridge: Cambridge University Press, 2011.
- [2] Coleman D Bernard, Noll Walter. Material symmetry and thermodynamic inequalities in finite elastic deformations[J]. Archive for Rational Mechanics and Analysis, 1964, 15(2): 87–111.
- [3] Dai Shuyang, Wang Fengru, Yang Jerry Zhijian, Yuan Cheng. On the cauchy-born approximation at finite temperature for alloys[J]. Discrete and Continuous Dynamical Systems-B, 2021, 27(6): 3131–3153.
- [4] Anthony Martin, Bartlett L Peter. Neural network learning: Theoretical foundations[M]. Cambridge: Cambridge University Press, 1999.
- [5] Hu Tianyi, Yang Jerry Zhijian, Yuan Cheng. Dpk: Deep neural network approximation of the first piola-kirchhoff stress[J]. Advances in Applied Mathematics and Mechanics, Submitted.