MOMENT TENSOR POTENTIALS: A CLASS OF SYSTEMATICALLY IMPROVABLE INTERATOMIC POTENTIALS[*]

ALEXANDER V. SHAPEEV[†]

Abstract. Density functional theory offers a very accurate way of computing materials properties from first principles. However, it is too expensive for modeling large-scale molecular systems whose properties are, in contrast, computed using interatomic potentials. The present paper considers, from a mathematical point of view, the problem of constructing interatomic potentials that approximate a given quantum-mechanical interaction model. In particular, a new class of systematically improvable potentials is proposed, analyzed, and tested on an existing quantum-mechanical database.

Key words. machine learning, interatomic potentials, moment tensor potentials

AMS subject classifications. 70G10, 68Q32

DOI. 10.1137/15M1054183

1. Introduction. Molecular modeling is an increasingly popular tool in biology, chemistry, physics, and materials science [15]. The success of molecular modeling largely depends on the accuracy and efficiency of calculating the interatomic forces. The two major approaches to computing interatomic forces are (1) quantum mechanics (QM) calculations [10, 14] and (2) using empirical interatomic potentials. In the first approach, first the electronic structure is computed and then the forces on the atoms (more precisely, on their nuclei) are deduced. QM calculations are therefore rather computationally demanding but can yield a high quantitative accuracy. On the other hand, the accuracy and transferability of empirical interatomic potentials is limited, but they are orders of magnitude less computationally demanding. With interatomic potentials, typically, the forces on atoms derive from the energy of interaction of each atom with its atomic neighborhood (typically consisting of tens to hundreds of atoms).

This hence makes it very attractive to design a combined approach whose efficiency would be comparable to interatomic potentials, yet the accuracy is similar to the one obtained with ab initio simulations. Moreover, the scaling of computational complexity of using the interatomic potentials is $O(N)$, where N is the number of atoms, whereas, for instance, standard Kohn–Sham density functional theory calculations scale like $O(N^3)$—which implies that the need in such combined approaches is even higher for large numbers of atoms.

One way of achieving this is designing accurate interatomic potentials. With this goal in mind, we categorize all potentials into two groups, following the regression analysis terminology of [3]. The first group is the parametric potentials with a fixed number of numerical or functional parameters. All empirical potentials known to the author are parametric, which is their disadvantage—their accuracy cannot be systematically improved. As a result, if the class of problems is sufficiently large and the required accuracy is sufficiently high (say, close to that of ab initio calculations), then the standard parametric potentials are not sufficient and one needs to employ nonparametric potentials. In theory, the accuracy of these potentials can be system-

atically improved. In practice, however, increasing the accuracy is done at a cost of higher computational complexity (which is nevertheless orders of magnitude less than that of the QM calculations), and the accuracy is still limited by that of the available QM model the potential is fitted to. Also, it should be noted that, typically, such fitted potentials do not perform well for the situations that they were not fitted to (i.e., they exhibit little or no transferability).

Each nonparametric potential has two major components: (1) a representation (also referred to as "descriptors") of atomic environments and (2) a regression model which is a function of the representation. It should be emphasized that finding a good representation of atomic environments is not a trivial task, as the representation is required to satisfy certain restrictions, namely, invariance with respect to Euclidean transformations (translations, rotations, and reflections) and permutation of chemically equivalent atoms, smoothness with respect to distant atoms coming to and leaving the boundary of the atomic environment, as well as completeness (for more details see the discussion in [3, 6, 13]). The latter restriction means that the representation should contain all the information to unambiguously reconstruct the atomic environment.

In this context, one common approach to constructing nonparametric potentials, called the neural networks potentials (NNP) [7, 8], is using artificial NNs as the regression model and a family of descriptors first proposed by Behler and Parrinello [8]. These descriptors have a form of radial functions applied to and summed over atomic distances to the central atom of the environment, augmented with similar functions with angular dependence summed over all pairs of atoms. Another approach, called Gaussian approximation potentials (GAP) [4, 27], uses Gaussian process regression and a cleverly designed set of descriptors that are evaluated by expanding smoothened density field of atoms into a spherical harmonics basis [3].

Other recent approaches include [28], which employs the bispectrum components of the atomic density, as proposed in an earlier version of GAP [4], and uses a linear regression model to fit the QM data. Also, [21] uses Gaussian process regression of a force-based model rather than the potential energy-based model. Additionally, in a recent work [19] the authors put forward an approach of regressing the interatomic interaction energy together with the electron density.

It is worthwhile to mention a related field of research, namely, the development of nonreactive interatomic potentials. Such potentials assume a fixed underlying atomistic structure (fixed interatomic bonds or fixed lattice sites that atoms are bound to). The works developing nonparametric versions of such potentials include [1, 17, 24].

In the present paper we propose a new approach to constructing nonparametric potentials based on linear regression and invariant polynomials. The main feature of this approach is that the proposed form of the potential can provably approximate any regular function satisfying all the needed symmetries (see Theorems 3.1 and 3.2), while being computationally efficient. The building block of the proposed potentials is what we call the moment tensors—similar to inertia tensors of atomic environments. Hence we call this approach the moment tensor potentials (MTP).

The manuscript is structured as follows. Section 2 gives an overview of interatomic potentials. Section 3 introduces MTP and then formulates and proves the two main theorems. The practical implementation of MTP is discussed in section 4. Section 5 reports the accuracy and computational efficiency (i.e., the CPU time) of MTP and compares MTP with GAP.

2. Interatomic potentials. Consider a system consisting of N_{tot} atoms with positions $x \in (\mathbb{R}^d)^{N_{\text{tot}}}$ (more precisely, with positions of their nuclei to be x), where d is the number of physical dimensions, taken as $d = 3$ in most of applications. These atoms have a certain energy of interaction, $E^{\mathrm{q}}(x)$, typically given by some QM model that we aim to approximate. For simplicity, we assume that all atoms are chemically equivalent.

We make the assumption that $E^{\mathrm{q}}(x)$ can be well approximated by a sum of energies of atomic environments of the individual atoms. We denote the atomic environment of atom k by Dx_k and let it equal to a tuple,

$$(2.1) \qquad Dx_k := (x_i - x_k)_{1 \leq i \leq N_{\text{tot}}, \, 0 < |x_i - x_k| \leq R_{\text{cut}}},$$

where R_{cut} is the cut-off radius, in practical calculations typically taken to be between 5 and 10Å. In other words, Dx_k is the collection of vectors joining atom k with all other atoms located at distance R_{cut} or closer. Thus, we are looking for an approximant of the form

$$(2.2) \qquad E(x) := \sum_{k=1}^{N_{\text{tot}}} V(Dx_k),$$

where V is called the interatomic potential. This assumption is true in most systems with short-range interactions (as opposed to, e.g., Coulomb interaction in charged or polarized systems); refer to recent works [11, 12, 23] for rigorous proofs of this statement for simple QM models.

Mathematically, since Dx_k can be a tuple of any size (in practice limited by the maximal density of atoms), V can be understood as a family of functions each having a different number of arguments. For convenience, however, we will still refer to V as a "function."

The function $V = V(u_1, \ldots, u_n)$ is required to satisfy the following restrictions:
(R1) Permutation invariance:

$$V(u_1, \ldots, u_n) = V(u_{\sigma_1}, \ldots, u_{\sigma_n}) \qquad \text{for any } \sigma \in \mathcal{S}_n,$$

where $\mathcal{S}_n$ denotes the set of permutations of $(1, \ldots, n)$.
(R2) Rotation and reflection invariance:

$$V(u_1, \ldots, u_n) = V(Qu_1, \ldots, Qu_n) \qquad \text{for any } Q \in O(d),$$

where $O(d)$ is the orthogonal group in $\mathbb{R}^d$. For simplicity in what follows we will denote $Qu := (Qu_1, \ldots, Qu_n)$.
(R3) Smoothness with respect to the number of atoms (more precisely, with respect to atoms leaving and entering the interaction neighborhood):

$$V : \mathbb{R}^{d \times n} \to \mathbb{R} \text{ is a smooth function,}$$
$$V(u_1, \ldots, u_n) = V(u_1, \ldots, u_n, u_{n+1}) \qquad \text{whenever } |u_{n+1}| \geq R_{\text{cut}}.$$

In most of the works performing practical calculations, including [6, 7, 8, 26, 27, 28], the interatomic potentials are chosen such that the energy and forces are continuous. There are, however, exceptions including [4] that require the second derivatives of energy to be continuous, and [20] proposing exponential decay of the potential with no strict cut-off radius.

Note that E is translation symmetric by definition.

2.1. Empirical interatomic potentials. For the purpose of illustration, we will briefly present two popular classes of empirical interatomic potentials. The first class is pair potentials (also known as two-body potentials),

$$V^{\mathrm{pair}}(u) = \sum_i \varphi(|u_i|),$$

where by u we denote the collection of the relative coordinates, $u = (u_i)_{i=1}^n$. It is clear that this potential satisfies (R1) and (R2), while if we take $\varphi(r) = 0$ for $r \geq R_{\mathrm{cut}}$ (as is done in most practical calculations), then it also satisfies (R3). The second potential is the embedded atom model,

$$V^{\mathrm{eam}}(u) = \sum_i \varphi(|u_i|) + F\Big(\sum_i \rho(|u_i|)\Big),$$

where if $\varphi(r)$ and $\rho(r)$ are chosen to vanish for $r \geq R_{\mathrm{cut}}$, then it also satisfies (R1)–(R3).

Both potentials can be viewed to have descriptors of the form

$$(2.3) \qquad R_\nu(u) = \sum_i f_\nu(|u_i|),$$

where f_ν are chosen such that their linear combination can approximate any smooth function (e.g., $f_\nu(r) = r^\nu(R_{\mathrm{cut}} - r)^2$ for $r \leq R_{\mathrm{cut}}$ and $f_\nu(r) = 0$ for $r > R_{\mathrm{cut}}$, where $\nu = 0, 1, \ldots$; the term $(R_{\mathrm{cut}} - r)^2$ ensures continuous energy and forces). Then, we can approximate $\varphi(r) \approx \sum_\nu c_\nu f_\nu(r)$ (the sum is taken over some finite set of ν), and hence

$$V^{\mathrm{pair}}(u) \approx \sum_\nu c_\nu R_\nu(u).$$

Likewise one can approximate V^{eam} with the exception that it will not be, in general, a linear function of $R_\nu(u)$.

This set of descriptors is not complete, because it is based only on the distances to the central atom, but is insensitive to bond angles.

2.2. Nonempirical interatomic potentials. Next, we review a number of nonempirical interatomic potentials proposed recently as a more accurate alternative to the empirical ones.

We start with the NNPs [6, 7, 8]. In addition to the descriptors given by (2.3) they also employ the descriptors of the form

$$(2.4) \qquad \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n f(|u_i|, |u_j|, u_i \cdot u_j).$$

(Note that NN can be, in principle, used with any set of descriptors. If these descriptors are complete, then such NNP is systematically improvable.) Typically, 50 to 100 descriptors of the form (2.3) and (2.4) are chosen as an input to an NN, while its output yields the function V [7]. In practice, this approach gives convincing results; however, it is an open question whether these descriptors are complete.

The GAP [4, 27] uses a different idea, consisting of (1) forming a smoothened atomic density function

$$\sum_{i=1}^n \exp\big(-\tfrac{|u_i - u|^2}{2\sigma^2}\big),$$

(2) approximating it through spherical harmonics, and (3) constructing functionals applied to the spherical harmonics coefficients that satisfy all needed symmetries. One can prove mathematically that this approach can approximate any regular symmetric function (i.e., a function that satisfies (R1)–(R3)). However, expanding functions in a spherical harmonics basis can be computationally expensive.

Several other nonempirical potentials have been proposed recently. A method using a representation based on expanding the atomic density function in spherical harmonics together with linear regression was used in [28]. In [21] the authors use Gaussian process regression to train a model that predicts forces on atoms directly (as opposed to a model that fits the energy).

3. Moment tensor potentials.

3.1. Representation with invariant polynomials. In the present paper we propose an alternative approach based on invariant polynomials. The idea is that any given potential $V^*(u)$, if it is smooth enough, can be approximated by a polynomial $p(u) \approx V^*(u)$—we will analyze the error of such an approximation in section 3.2. The approximant can always be chosen symmetric. Indeed, given $p(u)$, one can consider the symmetrized polynomial

$$(3.1) \qquad p^{\mathrm{sym}}(u_1, \ldots, u_n) = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} p(u_{\sigma_1}, \ldots, u_{\sigma_n}),$$

and then $V^*(u) \approx p^{\mathrm{sym}}(u)$. Similarly, one can symmetrize p^{sym} with respect to rotations and reflections. Hence, theoretically, one can construct a basis of such polynomials $b_\nu(u)$ and choose

$$V^*(u) \approx V(u) := \sum_\nu c_\nu b_\nu(u).$$

This approach is implemented for small systems of up to 10 atoms [9]; however, generalizations of this approach require a more efficient way of generating the invariant polynomials. The main difficulty is related to the fact that the number of permutations in a system of n atoms is $n!$, which grows too fast in order to, for instance, calculate the right-hand side of (3.1).

In the present paper we propose a basis for the set of all polynomials invariant with respect to permutation, rotation, and reflection. The main feature of the proposed basis is that the computational complexity of computing these polynomials scales like $O(n)$. Moreover, one can easily construct such bases to also satisfy the (R3) property (refer to section 4), making it a promising candidate for efficient nonparametric interatomic potentials.

The M polynomials. The building blocks of the basis functions for the approximation (representation) of $V = V(u)$ are the "moment" polynomials $M = M_{\bullet,\bullet}(u)$ defined as follows.

For integer $\mu, \nu \geq 0$ we let

$$M_{\mu,\nu}(u) := \sum_{i=1}^{n} |u_i|^{2\mu} u_i^{\otimes \nu},$$

where $w^{\otimes \nu} := w \otimes \cdots \otimes w$ is the Kronecker product of ν copies of the vector $w \in \mathbb{R}^d$. Thus, $M_{\mu,\nu}(u) \in (\mathbb{R}^d)^\nu$ (i.e., an ν-dimensional tensor) for each u. Computing $M_{\mu,\nu}(u)$

requires linear time in n, but exponential in ν. This means that if the maximal value of ν is bounded, then computing $M_{\mu,\nu}$ can be done efficiently.

There is a mechanical interpretation of $M_{\mu,\nu}(u)$. Consider $\mu = 0$; then $M_{0,0}$ simply gives the number of atoms at the distance of R_{cut} or less (this can also be understood as the "mass" of these atoms), $M_{0,1}$ is the center of mass of such atoms (scaled by the mass), $M_{0,2}$ is the tensor of second moments of inertia, etc. For $\mu > 0$, $M_{\mu,\nu}$ can be interpreted as weighted moments of inertia, with the ith atom's weight being $|u_i|^{2\mu}$.

The basis polynomials B_α. The basis polynomials are indexed by $k \in \mathbb{N}$, $k \le n$, where our definition for $\mathbb{N}$ is

$$\mathbb{N} = \{0, 1, \ldots\},$$

and a $k \times k$ symmetric matrix α of integers $\alpha_{i,j} \ge 0$ ($i, j \in \{1, \ldots, k\}$). For such matrices, by α_i' we define the sum of the off-diagonal elements of the ith row,

$$(3.2) \qquad \alpha_i' = \sum_{\substack{j=1 \\ j \ne i}}^{k} \alpha_{i,j}.$$

Next, we define a contraction operator (product) of tensors $T^{(i)} \in (\mathbb{R}^d)^{\alpha_i'}$ by

$$\prod_{i=1}^{\alpha \, k} T^{(i)} := \sum_{\beta} \prod_{i=1}^{k} T^{(i)}_{\beta^{(1,i)}\ldots\beta^{(i-1,i)}\beta^{(i,i+1)}\ldots\beta^{(i,k)}},$$

where each β is a collection of multi-indices $\beta = \left(\beta^{(i,j)}\right)_{1 \le i < j \le k}$, and each multi-index $\beta^{(i,j)}$ has the following form:

$$\beta^{(i,j)} = \left(\beta_1^{(i,j)}, \ldots, \beta_{\alpha_{i,j}}^{(i,j)}\right) \in \{1, \ldots, d\}^{\alpha_{i,j}}, \quad 1 \le i < j \le k.$$

To define this contraction rigorously, we let

$$(3.3) \quad \mathcal{M} := \{(i,j) \in \{1, \ldots, m\}^2 : i < j\}, \qquad \text{and} \qquad \mathcal{B} := (\{1, \ldots, d\}^{\alpha_{ij}})_{(i,j) \in \mathcal{M}}$$

and, expanding the multiindex notation, we write

$$\prod_{i=1}^{\alpha \, k} T^{(i)} = \sum_{\beta \in \mathcal{B}} \prod_{i=1}^{k} T^{(i)}_{\beta_1^{(1,i)}\ldots\beta_{\alpha_{1i}}^{(1,i)} \quad \ldots \quad \beta_1^{(i-1,i)}\ldots\beta_{\alpha_{i-1,i}}^{(i-1,i)} \ \beta_1^{(i,i+1)}\ldots\beta_{\alpha_{i,i+1}}^{(i,i+1)} \quad \ldots \quad \beta_1^{(i,k)}\ldots\beta_{\alpha_{i,k}}^{(i,k)}}.$$

Hence α_{ij} can be interpreted as how many dimensions are contracted between $T^{(i)}$ and $T^{(j)}$.

Finally, we let

$$(3.4) \qquad B_\alpha(u) := \prod_{i=1}^{\alpha \, k} M_{\alpha_{ii}, \alpha_i'}(u)$$

and call it a basis function (for a given α). Here $B_\alpha(u)$ is, essentially, a k-body function. We note that $B_\alpha(u) \equiv B_\beta(u)$ if there exists a permutation $\sigma \in \mathcal{S}_k$ such that $\alpha_{ij} = \beta_{\sigma_i \sigma_j}$ for all i and j.

For illustrative purposes we work out three examples of different B_α. First, let us take $\alpha = \left(\begin{smallmatrix} \mu_1 & 1 \\ 1 & \mu_2 \end{smallmatrix}\right)$. This implies $k = 2$ (since this is a 2×2 matrix) and $\alpha'_1 = \alpha'_2 = 1$. Hence B_α is the scalar product of $M_{\mu_1,1}$ and $M_{\mu_2,1}$; indeed,

$$B_\alpha = \sum_{\beta_1^{(1,2)}=1}^{d} \left(M_{\mu_1,1}\right)_{\beta_1^{(1,2)}} \left(M_{\mu_2,1}\right)_{\beta_1^{(1,2)}} = M_{\mu_1,1} \cdot M_{\mu_2,1},$$

where $\left(M_{\mu_1,1}\right)_{\beta_1^{(1,2)}}$ denotes the $\beta_1^{(1,2)}$th component of the vector $M_{\mu_1,1}$.

As the next example, we take $\alpha = \left(\begin{smallmatrix} \mu_1 & 2 \\ 2 & \mu_2 \end{smallmatrix}\right)$. Then $\alpha'_1 = \alpha'_2 = 2$, and hence B_α is the Frobenius product of the two matrices, $M_{\mu_1,2}$ and $M_{\mu_2,2}$:

$$B_\alpha = \sum_{\beta_1^{(1,2)}=1}^{d} \sum_{\beta_2^{(1,2)}=1}^{d} \left(M_{\mu_1,2}\right)_{\beta_1^{(1,2)},\beta_2^{(1,2)}} \left(M_{\mu_2,2}\right)_{\beta_1^{(1,2)},\beta_2^{(1,2)}} = M_{\mu_1,1} : M_{\mu_2,1}.$$

In the last example, we take $\alpha = \left(\begin{smallmatrix} \mu_1 & 1 & 1 \\ 1 & \mu_2 & 0 \\ 1 & 0 & \mu_3 \end{smallmatrix}\right)$. Then B_α is the following contraction of the matrix $M_{\mu_1,2}$ and two vectors, $M_{\mu_2,1}$ and $M_{\mu_2,1}$:

$$B_\alpha = \sum_{\beta_1^{(1,2)}=1}^{d} \sum_{\beta_1^{(1,3)}=1}^{d} \left(M_{\mu_1,2}\right)_{\beta_1^{(1,2)},\beta_1^{(1,3)}} \left(M_{\mu_2,1}\right)_{\beta_1^{(1,2)}} \left(M_{\mu_3,1}\right)_{\beta_1^{(1,3)}} = \left(M_{\mu_1,2} M_{\mu_2,1}\right) \cdot M_{\mu_3,1}.$$

Representability. We are now ready to formulate the main result, namely, that the linear combinations of B_α span all permutation and rotation invariant polynomials. Let us denote the set of all polynomials of n vector-valued variables by $\mathbb{P}$, the set of permutation invariant polynomials by $\mathbb{P}_{\text{perm}} \subset \mathbb{P}$, and the set of rotation invariant polynomials by $\mathbb{P}_{\text{rot}} \subset \mathbb{P}$.

THEOREM 3.1. *The polynomials B_α form a spanning set of the linear space $\mathbb{P}_{\text{rot}} \cap \mathbb{P}_{\text{perm}} \subset \mathbb{P}$ in the sense that any $p \in \mathbb{P}_{\text{rot}} \cap \mathbb{P}_{\text{perm}}$ can be represented by a (finite) linear combination of B_α (but this combination is, in general, not unique).*

We postpone the proof of this result to section 3.3, after we illustrate in section 3.2 that a QM model can be efficiently approximated with polynomials.

3.2. Approximation error estimate. In this section we present an error analysis of fitting the prototypical tight-binding QM model as proposed in [11, 12] with the polynomials $B_\alpha(u)$. We note that we directly take the site energy $V^{\text{q}}(u)$ constructed in [11, 12] (rather than starting from the total energy $E^{\text{q}}(x)$) and fix n to be constant throughout this analysis. The latter assumption needs some comment. Letting n be constant is essentially equivalent to assuming finite R_{cut}, provided that the atoms cannot come too close to each other. We note that making such assumption is not an insignificant limitation of the present analysis; however, we find lifting this limitation not at all trivial.

The QM model is defined as follows. We let, for convenience, $u_0 := 0$ and define the Hamiltonian matrix

$$H_{ij}(u) := \begin{cases} \varphi(u_j - u_i), & i,j \in \{0,\dots,n\} \text{ and } i \neq j, \\ 0, & i = j \in \{0,\dots,n\}, \end{cases}$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is an empirically chosen function referred to as the hopping integral [14].

Then we apply the function f^{q} to the matrix H (refer to [18] for details on matrix functions) and define V^{q} as the $(0,0)$th element of this matrix:

$$V^{\mathrm{q}}(u) := (f^{\mathrm{q}}(H))_{0,0},$$

where

$$f^{\mathrm{q}}(\epsilon) := \epsilon \left(1 + e^{\epsilon/(k_{\mathrm{B}}T)}\right)^{-1},$$

where $k_{\mathrm{B}} > 0$ is the Boltzmann constant, and $T > 0$ is the electronic temperature, and we take the chemical potential to be zero.

Next, we let $R > 0$ and $\delta_0 > 0$, allow u_i to vary in $\mathcal{V}_{\delta_0}$, where

$$\mathcal{V}_\delta := \{\zeta \in \mathbb{C} : \mathrm{Re}(\zeta) \in [-R-\delta, R+\delta]^d, \ \mathrm{Im}(\zeta) \in [-\delta, \delta]^d\},$$

and assume that $\varphi(v)$ is analytically extended to $\mathcal{V}_{\delta_0}$. We note that in many models $\varphi(\zeta)$ has a singularity at $\zeta = 0$ (for example, $\varphi(\zeta) = \beta_0 \exp(-q|\zeta|)$ for some β_0, $q > 0$ [14, equation (7.24)]); therefore by assuming analytical extensibility of φ onto $\mathcal{V}_{\delta_0}$ we implicitly assume some approximation of such irregular φ with a function that is regular around $\zeta = 0$. For example, one could use the Hermite functions basis, $e^{-|\zeta|^2/2}H_{n_1}(\zeta_1)H_{n_2}(\zeta_2)\dots H_{n_d}(\zeta_d)$ [25], to approximate $\varphi(\zeta)$ away from $\zeta = 0$.

THEOREM 3.2. *There exist $C > 0$ and $\rho > 1$, both depending only on n, δ_0, M_{δ_0}, and $k_{\mathrm{B}}T$, such that for any $m \in \mathbb{N}$ there exists $p_m \in \mathbb{P}_{\mathrm{perm}} \cap \mathbb{P}_{\mathrm{rot}}$ of degree m such that*

$$(3.5) \qquad \sup_{u:\, \max_i |u_i| \le R} |V^{\mathrm{q}}(u) - p_m(u)| < C\rho^{-m}.$$

The proof of this theorem is based on the following result by Hackbusch and Khoromskij [16].

PROPOSITION 3.3. *Let $\mathcal{E}_\rho := \{\zeta \in \mathbb{C} : |z-1| + |z+1| \le \rho + \rho^{-1}\}$, for some $\rho > 1$,*

$$\mathcal{E}_\rho^{(j)} := [-1,1]^{j-1} \times \mathcal{E}_\rho \times [-1,1]^{n-j-1},$$

$f = f(z_1, \dots, z_n)$ defined on the union of all $\mathcal{E}_\rho^{(j)}$, and

$$M_\rho(f) := \max_{1 \le j \le n} \sup_{z \in \mathcal{E}_\rho^{(j)}} |f(z)|.$$

Let $p = p(z_1, \dots, z_n)$ be the polynomial interpolant of f of degree m on the Chebyshev–Gauss–Lobatto nodes in $[-1,1]^n$. Then

$$\sup_{\max_i |z_i| \le 1} |f(z) - p(z)| \le 2(1 + 2\pi^{-1}\log(n))^m M_\rho(f)\frac{\rho^{-m}}{\rho - 1}.$$

An illustration of $\mathcal{E}_\rho$ and $\mathcal{V}_\delta$ is given in Figure 3.1.

Proof of Theorem 3.2. The plan of the proof is to, after an auxiliary step (Step 1), successively obtain bounds on φ (Step 2), on $H(u)$ (Step 3), and on the interpolating polynomial (Step 4), and then symmetrize this interpolating polynomial (Step 5).

Step 1. First, we define $M_\delta := \sup_{\zeta \in \mathcal{V}_\delta} \varphi(\zeta)$ and note that by the Cauchy integral formula one can bound

$$|\varphi'(z)| \le \left| \oint_{\partial \mathcal{V}_{2\delta}} \frac{\varphi(\zeta)\mathrm{d}\zeta}{(\zeta - z)^2} \right| = \frac{(2R + 8\delta)M_{2\delta}}{\delta^2} =: M'_\delta \qquad \forall z : |z| \le \delta.$$
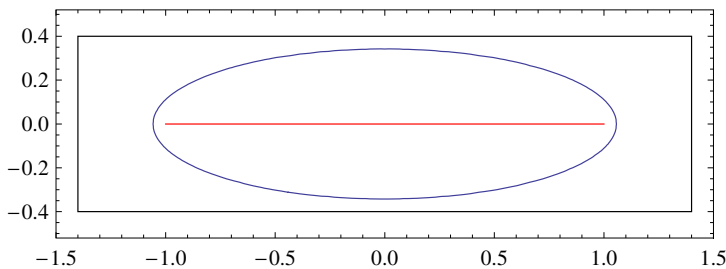
FIG. 3.1. *Illustration of $\mathcal{V}_\delta$ (black square) and $\mathcal{E}_\rho$ (blue ellipse) on the complex plane for $R = 1$, $\delta = 0.4$, $\rho = 1.4$. Both regions contain $[-1,1]$ (red line) and for this choice of parameters $\mathcal{E}_\rho \subset \mathcal{V}_\delta$.*

This is valid for any $\delta \leq \delta_0/2$.

Step 2. Next we let $\delta \in (0, \delta_0/2]$, which will be fixed later, and note that if $z \in \mathcal{V}_\delta$, then $|\mathrm{Im}(\varphi(z))| \leq \mathrm{Im}(\varphi(\mathrm{Re}(z))) + M'_{\delta_0}\mathrm{Im}(z) \leq M'_{\delta_0}\delta$ thanks to the intermediate value theorem.

Step 3. Next, following Proposition 3.3, define

$$\mathcal{U}_\delta := \cup_{i=1}^n \mathcal{V}_0^{i-1} \times \mathcal{V}_\delta \times \mathcal{V}_0^{n-i-1}.$$

Hence note that for $u \in \mathcal{U}_\delta$, $H(u)$ is symmetric (possibly non-Hermitian) with at most $2n$ elements being nonreal, which makes it easy to estimate using the Frobenius norm:

$$\|\mathrm{Im}(H(u))\| \leq \sqrt{2n}M'_\delta\delta.$$

By a spectrum perturbation argument (namely, the Bauer–Fike theorem [5]) we have the corresponding bound on the spectrum:

$$|\mathrm{Im}(\mathrm{Sp}(H(u)))| \leq \sqrt{2n}M'_\delta\delta.$$

The real part of the spectrum can be estimated directly through the norm:

$$|\mathrm{Re}(\mathrm{Sp}(H(u)))| \leq \|H(u)\| \leq nM_{\delta_0}.$$

Step 4. We next bound $V^{\mathrm{q}}(u)$ for $u \in \mathcal{U}_\delta$. If needed we decrease δ such that $\sqrt{2n}M'_\delta\delta < \frac{\pi}{3}k_{\mathrm{B}}T$ and use the following representation [11, 18]:

$$V^{\mathrm{q}}(u) = -\frac{1}{2\pi\mathrm{i}} \oint_\gamma f^{\mathrm{q}}(z)\big((H(u) - zI)^{-1}\big)_{0,0}\mathrm{d}z,$$

where we take $\gamma := \partial\Omega$, where $\Omega = \{z \in \mathbb{C} : |\mathrm{Re}(z)| \leq nM_{\delta_0} + \frac{\pi}{3}k_{\mathrm{B}}T, |\mathrm{Im}(z)| \leq \frac{2\pi}{3}k_{\mathrm{B}}T\}$. The choice of the region Ω is such that any point $z \in \gamma$ is separated from any eigenvalue λ of $H(u)$ by the distance $|z - \lambda| \geq \frac{\pi}{3}k_{\mathrm{B}}T$, and at the same time $f^{\mathrm{q}}(z)$ is regular on $\overline{\Omega}$. This allows us to estimate, for $z \in \gamma$,

$$\|(H(u) - zI)^{-1}\| \leq \big(\tfrac{\pi}{3}k_{\mathrm{B}}T\big)^{-1}$$

and hence

$$\sup_{u \in \mathcal{U}_\delta} |V^{\mathrm{q}}(u)| \leq \sup_{z \in \Omega} |f^{\mathrm{q}}(z)| |\gamma| \big(\tfrac{\pi}{3}k_{\mathrm{B}}T\big)^{-1} \leq C,$$

where C is a generic constant that depends only on n, δ_0, M_{δ_0}, and $k_{\mathrm{B}}T$. It now remains to notice that there exists $\rho > 1$ that depends on δ such that

$$\{z \in \mathbb{C} : |z/R - 1| + |z/R + 1| \leq \rho + \rho^{-1}\} \subset \mathcal{V}_\delta.$$

Hence Proposition 3.3 applies to the function $f(z) = V^{\mathrm{q}}(Rz)$ and yields the interpolating polynomial $\tilde{p}_m(u)$ such that

$$\sup_{u:\,\max_i |u_i|\leq R} |V^{\mathrm{q}}(u) - \tilde{p}_m(u)| \leq \sup_{u:\,\max_i |u_i|_\infty\leq R} |V^{\mathrm{q}}(u) - \tilde{p}_m(u)| \leq C\rho^{-m}.$$

By construction $\tilde{p}_m \in \mathbb{P}_{\mathrm{perm}}$. Indeed, the function $f(z) = V^{\mathrm{q}}(Rz)$ is symmetric with respect to permutation of variables, and so is the Chebyshev–Gauss–Lobatto interpolation nodes on the domain $[-1,1]^n$. Hence, uniqueness of interpolation yields permutation symmetry of $\tilde{p}_m$.

Step 5. We define

$$p_m(u) := \frac{\int_{Q\in O(d)} \tilde{p}_m(Qu)\mathrm{d}Q}{\int_{Q\in O(d)} \mathrm{d}Q} \in \mathbb{P}_{\mathrm{perm}} \cap \mathbb{P}_{\mathrm{rot}}$$

(where $\mathrm{d}Q$ denotes the Haar measure) and thanks to the rotation invariance of V^{q} we recover (3.5). □

It is worth noting that the integration with respect to rotation was used only as a technical tool in the proof. If we directly constructed an approximation $V(u) = \sum_\alpha c_\alpha B_\alpha(u)$ to $V^{\mathrm{q}}(u)$, rather than using the Chebyshev–Gauss–Lobatto interpolation as an intermediate step, then $V(u)$ would be rotationally invariant by construction.

3.3. Proof of Theorem 3.1. We start with stating the first fundamental theorem for the orthogonal group $O(d)$ [30].

THEOREM 3.4. *$p \in \mathbb{P}$ is rotation invariant if and only if it can be represented as a polynomial of $n(n + 1)/2$ scalar variables of the form $r_{ij}(u) := u_i \cdot u_j$, where $1 \leq i \leq j \leq n$.*

We hence can identify a polynomial $p = p(u) \in \mathbb{P}_{\mathrm{rot}}$ with the respective polynomial $q = q(r) \in \mathbb{Q}$, where $\mathbb{Q}$ is the set of polynomials of $n(n + 1)$ scalar variables $r = (r_{ij})_{1\leq i,j\leq n}$ (we lift the requirement $i \leq j$ for ease of notation in what follows).

In order to proceed, we introduce the notation of composition of tuples:
(3.6)
$$a_b := (a_{b_1}, \ldots, a_{b_m}) \qquad \text{for} \ \ b = (b_1, \ldots, b_m) \subset \{1, \ldots, n\}^m \ \ \text{and} \ \ a = (a_1, \ldots, a_n).$$

Hence define

$$\mathbb{Q}_{\mathrm{perm}} := \{q \in \mathbb{Q} : q(r) \equiv q(r_{\sigma\sigma}) \ \forall \sigma \in \mathcal{S}_n\}$$

that corresponds to $\mathbb{P}_{\mathrm{perm}}$, where we likewise let $r_{\sigma\sigma} := (r_{\sigma_i\sigma_j})_{1\leq i,j\leq n}$.

We next formulate a rather intuitive result, essentially stating that $\mathbb{Q}_{\mathrm{perm}}$ can be spanned by symmetrizing all monomials of r.

LEMMA 3.5.

$$\mathbb{Q}_{\mathrm{perm}} = \mathrm{Span}\left\{ \sum_{\sigma\in\mathcal{S}_n} \prod_{i=1}^m \prod_{j=i}^m (r_{\sigma_i\sigma_j})^{\alpha_{ij}} : m \in \mathbb{N}, \ m \leq n, \ \alpha \in \mathbb{N}^{m\times m}\right\}.$$

Proof. If $q \in \mathbb{Q}_{\mathrm{perm}}$, then

$$\frac{1}{n!} \sum_{\sigma\in\mathcal{S}_n} q(r_{\sigma\sigma}) = \frac{1}{n!} \sum_{\sigma\in\mathcal{S}_n} q(r) = q(r).$$

It remains to apply this identity to all monomials $q(r) = \prod_{i=1}^m \prod_{j=i}^m r_{ij}^{\alpha_{ij}}$ to obtain the stated result. □

The next step toward proving the main result is the following lemma, where we denote $\mathcal{N} := \{1, \dots, n\}$.

LEMMA 3.6. *For $m \in \mathbb{N}$, $m \le n$, and $\alpha \in \mathbb{N}^{m \times m}$,*

$$B_\alpha(u) = \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} \prod_{j=i}^{m} (u_{\gamma_i} \cdot u_{\gamma_j})^{\alpha_{ij}}.$$

Proof. Before commencing with the proof, we note that we will use the distributive law of addition and multiplication in the following form:

$$\prod_{j=1}^{m} \sum_{i \in \mathcal{I}} f(i,j) = \sum_{\gamma \in \mathcal{I}^m} \prod_{j=1}^{m} f(\gamma_j, j).$$

Then

$$T^{(i)}_{\beta^{(1,i)} \dots \beta^{(i-1,i)} \beta^{(i,i+1)} \dots \beta^{(i,m)}}(u) = \sum_{\gamma \in \mathcal{N}} |u_\gamma|^{2\alpha_{ii}} \left(\prod_{j=1}^{i-1} \left(u_\gamma^{\otimes \alpha_{ji}} \right)_{\beta^{(j,i)}} \right) \left(\prod_{j=i+1}^{m} \left(u_\gamma^{\otimes \alpha_{ij}} \right)_{\beta^{(i,j)}} \right)$$

$$= \sum_{\gamma \in \mathcal{N}} |u_\gamma|^{2\alpha_{ii}} \left(\prod_{j=1}^{i-1} \prod_{\ell=1}^{\alpha_{ji}} u_{\gamma, \beta_\ell^{(j,i)}} \right) \left(\prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} u_{\gamma, \beta_\ell^{(i,j)}} \right).$$

We recall the notation α_i' and $\mathcal{B}$, introduced in (3.2) and (3.3), respectively, and hence express

$$^\alpha \prod_{i=1}^{m} M_{\alpha_{ii}, \alpha_i'}(u) = \sum_{\beta \in \mathcal{B}} \prod_{i=1}^{m} T^{(i)}_{\beta^{(1,i)} \dots \beta^{(i-1,i)} \beta^{(i,i+1)} \dots \beta^{(i,m)}}$$

$$= \sum_{\beta \in \mathcal{B}} \prod_{i=1}^{m} \sum_{\gamma \in \mathcal{N}} |u_\gamma|^{2\alpha_{ii}} \left(\prod_{j=1}^{i-1} \prod_{\ell=1}^{\alpha_{ji}} u_{\gamma, \beta_\ell^{(j,i)}} \right) \left(\prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} u_{\gamma, \beta_\ell^{(i,j)}} \right)$$

$$= \sum_{\beta \in \mathcal{B}} \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \left(\prod_{j=1}^{i-1} \prod_{\ell=1}^{\alpha_{ji}} u_{\gamma_i, \beta_\ell^{(j,i)}} \right) \left(\prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} u_{\gamma_i, \beta_\ell^{(i,j)}} \right)$$

$$= \sum_{\beta \in \mathcal{B}} \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} \left(u_{\gamma_i, \beta_\ell^{(i,j)}} \right) \left(u_{\gamma_j, \beta_\ell^{(i,j)}} \right)$$

$$= \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \prod_{j=i+1}^{m} \sum_{\beta \in \{1, \dots, d\}^{\alpha_{ij}}} \prod_{\ell=1}^{\alpha_{ij}} \left(u_{\gamma_i, \beta_\ell} \right) \left(u_{\gamma_j, \beta_\ell} \right)$$

$$= \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} \sum_{\beta=1}^{d} (u_{\gamma_i, \beta})(u_{\gamma_j, \beta})$$

$$= \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \prod_{j=i+1}^{m} \prod_{\ell=1}^{\alpha_{ij}} u_{\gamma_i} \cdot u_{\gamma_j}$$

$$= \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} |u_{\gamma_i}|^{2\alpha_{ii}} \prod_{j=i+1}^{m} \left(u_{\gamma_i} \cdot u_{\gamma_j} \right)^{\alpha_{ij}}$$

$$= \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m} \prod_{j=i}^{m} \left(u_{\gamma_i} \cdot u_{\gamma_j} \right)^{\alpha_{ij}}. \qquad \square$$

Proof of Theorem 3.1. In view of the previous lemma, we can denote

$$\tilde{\mathbb{Q}}_{\mathrm{perm}} = \mathrm{Span}\left\{ \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\gamma_i \gamma_j})^{\alpha_{ij}} : m \in \mathbb{N},\ m \leq n,\ \alpha \in \mathbb{N}^{m \times m} \right\}$$

and, applying the earlier lemmas, formulate the statement of Theorem 3.1 as $\tilde{\mathbb{Q}}_{\mathrm{perm}} = \mathbb{Q}_{\mathrm{perm}}$. The latter is an immediate corollary of the following more specialized result. □

LEMMA 3.7. *For* $m \in \mathbb{N}$, $m \leq n$, *denote*

$$\mathbb{Q}_{\mathrm{perm}}^{(m)} := \mathrm{Span}\left\{ \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\sigma_i \sigma_j})^{\alpha_{ij}} : \alpha \in \mathbb{N}^{m \times m} \right\} \qquad and$$

$$\tilde{\mathbb{Q}}_{\mathrm{perm}}^{(m)} := \mathrm{Span}\left\{ \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\gamma_i \gamma_j})^{\alpha_{ij}} : \alpha \in \mathbb{N}^{m \times m} \right\}.$$

Then $\tilde{\mathbb{Q}}_{\mathrm{perm}}^{(m)} = \mathbb{Q}_{\mathrm{perm}}^{(m)}$.

Before we prove this lemma, we give two auxiliary results.

LEMMA 3.8. *We equip* $\mathcal{N}^m$ *with the lexicographical order and hence denote by* $\Gamma := \{\gamma \in \mathcal{N}^m : \gamma = \min\{\sigma_\gamma : \sigma \in \mathcal{S}_n\}\}$ *the set of representatives of equivalence classes* $\{\sigma_\gamma : \sigma \in \mathcal{S}_n\}$, *where* σ_γ *is understood as composition of tuples* (3.6)*. Also let* $C_\gamma = \#\{\sigma_\gamma : \sigma \in \mathcal{S}_n\}$, *where* # *denotes the number of elements in a set. Then*

$$\sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\gamma_i \gamma_j})^{\alpha_{ij}} = \sum_{\gamma \in \Gamma} \frac{C_\gamma}{n!} \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\sigma_{\gamma_i} \sigma_{\gamma_j}})^{\alpha_{ij}}.$$

Proof. Any $\sigma \in \mathcal{S}_n$ induces a one-to-one mapping $\gamma \mapsto \sigma_\gamma$ on $\mathcal{N}^m$. Hence

$$\sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\gamma_i \gamma_j})^{\alpha_{ij}} = \sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\sigma_{\gamma_i} \sigma_{\gamma_j}})^{\alpha_{ij}} \qquad \forall \sigma \in \mathcal{S}_n,$$

therefore

$$\sum_{\gamma \in \mathcal{N}^m} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\gamma_i \gamma_j})^{\alpha_{ij}} = \sum_{\gamma \in \mathcal{N}^m} \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{m}\prod_{j=i}^{m} (r_{\sigma_{\gamma_i} \sigma_{\gamma_j}})^{\alpha_{ij}}.$$

It remains to group up the terms for which σ_γ is the same. □

The next auxiliary result is proved by a trivial combinatorial argument, essentially expressing that all elements of Γ other than $(1, \ldots, m) \in \Gamma$ have repeated values.

PROPOSITION 3.9. *Let* $m \geq 1$. *Then* $\Gamma = \{(1, \ldots, m)\} \cup \Gamma'$, *where* $\Gamma' := \{\gamma \in \Gamma : \max_i \gamma_i \leq m - 1\}$.

Proof of Lemma 3.7. We argue by induction over m. For $m = 0$ the statement is obvious since $\mathbb{Q}_{\mathrm{perm}}^{(0)} = \tilde{\mathbb{Q}}_{\mathrm{perm}}^{(0)} = \mathrm{Span}\{1\}$; therefore we only need to prove the induction step.

Now, for $m \in \mathbb{N}$, $m \leq n$, we choose an arbitrary α and let

$$q(r) := \sum_{\sigma \in S_n} \prod_{j,k=1}^{m} (r_{\sigma_j \sigma_k})^{\alpha_{jk}} \in \mathbb{Q}_{\mathrm{perm}} \qquad and$$

$$\tilde{q}(r) := \frac{n!}{C_{\{1,\ldots,m\}}} \sum_{\gamma \in \mathcal{N}^m} \prod_{j,k=1}^{m} (r_{\gamma_j \gamma_k})^{\alpha_{jk}} \in \tilde{\mathbb{Q}}_{\mathrm{perm}}.$$

These $q(r)$ and $\tilde{q}(r)$ span $\mathbb{Q}_{\text{perm}}^{(m)}$ and $\tilde{\mathbb{Q}}_{\text{perm}}^{(m)}$, respectively. We aim to show that $q(r) - \tilde{q}(r) \in \mathbb{Q}_{\text{perm}}^{(m-1)} = \tilde{\mathbb{Q}}_{\text{perm}}^{(m-1)}$. This will prove the result considering that, by definition, $\mathbb{Q}_{\text{perm}}^{(m-1)} \subseteq \mathbb{Q}_{\text{perm}}^{(m)}$ and $\tilde{\mathbb{Q}}_{\text{perm}}^{(m-1)} \subseteq \tilde{\mathbb{Q}}_{\text{perm}}^{(m)}$.

Indeed, in view of the two previous results, recall the definition $\Gamma' := \{\gamma \in \Gamma : \max_i \gamma_i \leq m - 1\}$ and write

$$\tilde{q}(r) - q(r) = \frac{n!}{C_{\{1,\ldots,m\}}} \sum_{\gamma \in \Gamma} \frac{C_\gamma}{n!} \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{m} \prod_{j=i}^{m} (r_{\sigma_{\gamma_i} \sigma_{\gamma_j}})^{\alpha_{ij}} - \sum_{\sigma \in S_n} \prod_{i=1}^{m} \prod_{j=i}^{m} (r_{\sigma_i \sigma_j})^{\alpha_{ij}}$$

$$= \sum_{\gamma \in \Gamma'} \frac{C_\gamma}{C_{\{1,\ldots,m\}}} \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{m} \prod_{j=i}^{m} (r_{\sigma_{\gamma_i} \sigma_{\gamma_j}})^{\alpha_{ij}}.$$

Next, we denote

$$\alpha_{k\ell}^{(\gamma)} := \sum_{\substack{1 \leq i \leq j \leq m \\ \gamma_i = k, \ \gamma_j = \ell}} \alpha_{ij}$$

and hence express

$$\tilde{q}(r) - q(r) = \sum_{\gamma \in \Gamma'} \frac{C_\gamma}{C_{\{1,\ldots,m\}}} \sum_{\sigma \in \mathcal{S}_n} \prod_{k=1}^{m-1} \prod_{\ell=i}^{m-1} (r_{\sigma_k \sigma_\ell})^{\alpha_{k\ell}^{(\gamma)}}.$$

Note that the upper limit of both products is $m - 1$ thanks to the definition of Γ' (and, of course, Proposition 3.9). Also, note that we used the fact that, from the definition of $\alpha^{(\gamma)}$, $\alpha_{k\ell}^{(\gamma)} = 0$ whenever $k > \ell$. Since $\sum_{\sigma \in \mathcal{S}_n} \prod_{k=1}^{m-1} \prod_{\ell=i}^{m-1} (r_{\sigma_k \sigma_\ell})^{\alpha_{k\ell}^{(\gamma)}}$ by definition belongs to $\mathbb{Q}_{\text{perm}}^{(m-1)}$ for each γ, we finally derive that

$$\tilde{q}(r) - q(r) \in \mathbb{Q}_{\text{perm}}^{(m-1)} = \tilde{\mathbb{Q}}_{\text{perm}}^{(m-1)}.$$

This concludes the proof of the induction step. $\square$

4. Practical implementation. The representation of the interatomic potential through polynomials, outlined in section 3.1, does not satisfy the (R3) property needed for the practical implementation. Hence, as the next step we modify the interatomic potential to satisfy this property. After this, in section 4.1 we discuss the steps needed to compute $B_\alpha(u)$, and in section 4.2 we describe the algorithms we used for fitting the potentials.

First, notice that for a fixed ν, a linear combination of moment tensors $M_{\mu,\nu}(u)$ is a polynomial of $|u_i|^2$ multiplied by $u_i^{\otimes \nu}$. The space of polynomials of $|u_i|^2$ can be substituted with any other space of functions (which, for generality, can be made dependent on ν), provided that they can represent any regular function of $|u|$, i.e.,

$$(4.1) \qquad \tilde{M}_{\mu,\nu}(u) := \sum_{i=1}^{n} f_{\mu,\nu}(|u_i|) u_i^{\otimes \nu},$$

where, e.g., $f_{\mu,\nu}(r) = r^{-\mu-\nu} f_{\text{cut}}(r)$ or $f_\mu(r) = e^{-k_\mu r} f_{\text{cut}}(r)$, $k_\mu > 0$ is some sequence of real numbers, and $f_{\text{cut}}(r)$ is some cut-off function such that $f_{\text{cut}}(r) = 0$ for $r \geq R_{\text{cut}}$. Here $f_{\mu,\nu}(r)$ plays essentially the same role as the "radial symmetry functions" in the NNP [7] or the radial basis functions in [3]. We then let

$$\tilde{B}_\alpha(u) := {}^\alpha\prod_{i=1}^{k} \tilde{M}_{\alpha_{ii}, \alpha_i'}(u)$$

(cf. (3.4)) and define the interatomic potential by

$$V(u) = \sum_{\alpha \in A} c_\alpha \tilde{B}_\alpha(u),$$

where A is a set of matrix-valued indices fixed a priori and c_α is the set of coefficients found in the training stage. For the rest of the paper we will omit tildes in $\tilde{M}_{\bullet,\bullet}$ and $\tilde{B}_\bullet$.

4.1. Computing the energy and forces. Next, we discuss the steps needed to compute the interatomic potential and its derivatives for a given neighborhood u. The computation consists of two parts, the precomputation (offline) step and the evaluation (online) step. The precomputation step accepts the set A of values of α as an input and generates the data for the next step, which is the efficient calculation of $B_\alpha(u)$ for a given neighborhood u.

Before we proceed, we make two observations.

1. The elements of the tensors $M_{\mu,\nu}$ are the "moments" of the form

$$m_{\mu,\beta}(u) := \sum_{i=1}^{n} f_{\mu,\nu}(|u_i|) \prod_{j=1}^{d} u_{i,j}^{\beta_j},$$

where β is a multi-index such that $|\beta| = \nu$. The higher the dimension is, the more repeated moments each $M_{\mu,\nu}$ contains (e.g., each matrix $M_{\mu,2}$ has nine elements, out of which at most six may be different due to the symmetricity of $M_{\mu,2}$).

2. The scalar functions B_α consist of products of $M_{\mu,\nu}$ (which means that the elements of B_α are linear combinations of products of $m_{\mu,\beta}$). Differentiating products of two terms is easier than differentiating products of three or more terms. Hence it will be helpful to have a representation of B_α as a product of two tensors.

To that end, we extend the definition of the product $\prod^\alpha$ by allowing the result to be a tensor of nonzero dimension:

$$\left({}^\alpha\prod_{i=1}^{k} T^{(i)} \right)_{\beta^{(1,1)}...\beta^{(k,k)}} := \sum_{\beta \in \mathcal{B}} \prod_{i=1}^{k} T^{(i)}_{\beta^{(i,1)}...\beta^{(i,k)}},$$

where each $T^{(i)}$ is a tensor of dimension $\sum_{j=1}^{k} \alpha_{ij}$ (here we let $\alpha_{ij} = \alpha_{ji}$), $\mathcal{B}$ is defined in (3.3), and we make a convention that $\beta^{(i,j)} := \beta^{(j,i)}$ for $i > j$. Hence we define

$$\hat{B}_{\bar\alpha,\alpha}(u) := {}^\alpha\prod_{i=1}^{k} M_{\bar\alpha_i, \sum_{j=1}^{k} \alpha_{ij}}(u),$$

parametrized by $k \in \mathbb{N}$, $\bar\alpha \in \mathbb{N}^k$, and a symmetric matrix $\alpha \in \mathbb{N}^{k \times k}$. $\hat{B}_{\bar\alpha,\alpha}(u)$ is a tensor of dimension $\sum_{i=1}^{k} \alpha_{ii}$. Clearly if $\bar\alpha_i = \alpha_{ii}$ for all i, then $\hat{B}_{\bar\alpha,\alpha} = B_\alpha$, which makes the collection of tensors $\hat{B}_{\bar\alpha,\alpha}$ a generalization of B_α.

Next, consider $\hat{B}_{\bar\alpha,\alpha}(u)$ for some $\bar\alpha \in \mathbb{N}^k$ and $\alpha \in \mathbb{N}^{k \times k}$, fix $1 \leq I \leq k$, and denote $\bar\beta = (\bar\alpha_1, \ldots, \bar\alpha_I)$, $\bar\gamma = (\bar\alpha_{I+1}, \ldots, \bar\alpha_k)$,

$$\beta := \begin{pmatrix} \alpha_{11} + \sum_{i=I+1}^{k} \alpha_{i1} & \alpha_{12} & \cdots & \alpha_{1I} \\ \alpha_{12} & \alpha_{22} + \sum_{i=I+1}^{k} \alpha_{i2} & \cdots & \alpha_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1I} & \alpha_{2I} & \cdots & \alpha_{II} + \sum_{i=I+1}^{k} \alpha_{iI} \end{pmatrix},$$

and

$$\gamma := \begin{pmatrix} \alpha_{I+1,I+1} + \sum_{i=1}^{I} \alpha_{i,I+1} & \alpha_{I+1,I+2} & \cdots & \alpha_{I+1,k} \\ \alpha_{I+1,I+2} & \alpha_{I+2,I+2} + \sum_{i=1}^{I} \alpha_{i,I+2} & \cdots & \alpha_{I+2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{I+1,k} & \alpha_{I+2,k} & \cdots & \alpha_{kk} + \sum_{i=1}^{I} \alpha_{i,k} \end{pmatrix}.$$

One can then express the elements of $\hat{B}_{\bar{\alpha},\alpha}(u)$ through products of elements of $\hat{B}_{\bar{\beta},\beta}(u)$ and $\hat{B}_{\bar{\gamma},\gamma}(u)$. Note that by reordering the rows and columns of α and β, one can generate many different ways of representing $\hat{B}_{\bar{\alpha},\alpha}(u)$. When doing computations, one should exercise this freedom in such a way that the resulting tensors are of minimal dimension so that the total computation cost is reduced.

Finally, note that even if $\hat{B}_{\bar{\alpha},\alpha}(u)$ was scalar-valued, $\hat{B}_{\bar{\beta},\beta}(u)$ and $\hat{B}_{\bar{\gamma},\gamma}(u)$ do not have to be scalar-valued—this motivates the need to introduce the tensor-valued basis functions $\hat{B}_{\bar{\alpha},\alpha}(u)$.

Precomputation. The precomputation step is hence as follows (we keep the argument u of, e.g., in $B_{\alpha}(u)$ to match the above notation; however, the particular values of u never enter the precomputation step):

- Starting with the set of tensors $B_{\alpha}(u)$ (indexed by $\alpha \in A$), establish their correspondence to $\hat{B}_{\bar{\alpha},\alpha}(u)$ and recursively represent each $\hat{B}_{\bar{\alpha},\alpha}(u)$ through some $\hat{B}_{\bar{\beta},\beta}(u)$ and $\hat{B}_{\bar{\gamma},\gamma}(u)$ as described above. In each case, out of all such products choose the one with the minimal sum of the number of dimensions of these two tensors.
- Enumerate all the elements of all the tensors $\hat{B}_{\bar{\alpha},\alpha}(u)$ as $b_i(u)$ (i is the ordinal number of the corresponding element).
- Represent each $b_i(u)$ as either $b_i(u) = m_{\mu_i,\beta_i}(u)$ or $b_i(u) = \sum_{j=1}^{J_i} c_j b_{\ell_j}(u) b_{k_j}(u)$.
- Output the resulting
  - (1) correspondence of $B_{\alpha}(u)$ to $b_i(u)$,
  - (2) μ_i and β_i, and
  - (3) tuples of (i, c, ℓ, k) corresponding to $b_i(u) = \sum_{j=1}^{J_i} c_j b_{\ell_j}(u) b_{k_j}(u)$.

Evaluation. The evaluation step, as written out below, accepts u as an input and evaluates $B_{\alpha}(u)$ using the precomputed data (described above).

1. For a given u, calculate all $m_{\mu_i,\beta_i}(u)$.
2. Then calculate all other $b_i(u)$ using the tuples of (i, c, ℓ, k).
3. Finally, pick those $b_i(u)$ that correspond to scalar $B_{\alpha}(u)$ (as opposed to non-scalar $\hat{B}_{\bar{\alpha},\alpha}$).

It remains to form the linear combination of $B_{\alpha}(u)$ with the coefficients obtained from a linear regression (training), and then sum up all these linear combinations for all atomic environments to obtain the interatomic interaction energy of a given atomistic system. The forces are computed by reverse-mode differentiation of the energy with respect to the atomic positions [22].

4.2. Training. Once the set A of values of α is fixed, we need to determine the coefficients c_{α}. This is done with the regularized (to avoid overfitting [2]) linear regression in the following way.

Let a database of atomic configurations $X = \{x^{(k)} : k = 1, \ldots, K\}$, where $x^{(k)}$ is of size $N^{(k)}$, be given together with their reference energies and forces, $E^{q}(x^{(k)}) = E^{(k)}$ and $-\nabla E^{q}(x^{(k)}) = f^{(k)}$. We form an overdetermined system of linear equations on c_{α},

$$\sum_{i=1}^{N^{(k)}} \sum_{\alpha \in A} c_\alpha B(Dx_i^{(k)}) = E^{(k)},$$

$$\frac{\partial}{\partial x_j^{(k)}} \sum_{i=1}^{N^{(k)}} \sum_{\alpha \in A} c_\alpha B(Dx_i^{(k)}) = -f_j^{(k)}$$

(cf. (2.1) and (2.2)), which we write in the matrix form $Xc = g$. These equations may be ill-conditioned, hence a regularization must be used. We tried three versions of regularization, namely, the ℓ_p regularization with $p \in \{0, 1, 2\}$, all described below. All three can be written as

$$\text{find} \quad \min_c \|Xc - g\|^2 \quad \text{subject to} \quad \|c\|_{\ell_p}^2 \le t,$$

where t is the regularization parameter and $\|c\|_{\ell_0}$ is defined as the number of nonzero entires in c. For $p \ge 1$ this can be equivalently rewritten as

$$\text{find} \quad \min_c \|Xc - g\|^2 + \gamma \|c\|_{\ell_p}^2,$$

where γ is an alternative regularization parameter.

4.2.1. ℓ_2 regularization. For the solution of the overdetermined linear equations $Xc = g$ we take

$$c = (X^T X + \gamma \operatorname{diag}(X^T X))^{-1} X^T g,$$

where $\operatorname{diag}(A)$ denotes the diagonal matrix whose diagonal elements are the same as in A. The penalization matrix was chosen as $\operatorname{diag}(X^T X)$ instead of the identity matrix so that its scaling with respect to the database size and the scale of the basis functions B_α, is compatible with that of the covariance matrix $X^T X$. The regularization parameter γ was determined from the 16-fold cross-validation scheme [2] as described in the next paragraph.

To perform the 16-fold cross-validation, we split the database X evenly into 16 nonoverlapping databases $\tilde{X}_1, \ldots, \tilde{X}_{16}$. We then train 16 different models, each on the database $X \setminus \tilde{X}_i$, and find the root mean square (RMS) error when tested on $\tilde{X}_i$. The parameter γ is then chosen such that the cross-validation RMS error averaged over these 16 models is minimal.

4.2.2. ℓ_0 regularization. The advantage of the ℓ_0 regularization is that it produces sparse solutions c, whereas the ℓ_2 regularization does not. We note that the ℓ_1 regularization also produces sparse solutions [29], but our numerical experiments show that the ℓ_0 regularization produces significantly more sparse solutions (however, at a cost of a larger precomputation time).

We thus solve a sequence of problems, parametrized by an integer parameter N_{nz} (number of nonzeros) as follows:

$$\text{find} \min_c \|Xc - g\| \text{ subject to } \|c\|_{\ell_0} = N_{\text{nz}},$$

and choose the minimal N_{nz} such that $\|Xc - g\|$ reaches the accuracy goal.

To describe the algorithm, it is convenient to rewrite the problem as follows. Let $\boldsymbol{A}$ be the set of all indices α (earlier denoted as A).

$$\text{find } A \subset \boldsymbol{A} \text{ such that } \min_c \|Xc - g\| \text{ is minimal}$$

$$\text{subject to } |A| = N_{\text{nz}} \text{ and } c_{\boldsymbol{A} \setminus A} = 0.$$

This is essentially a compressed sensing problem [31]. In order to solve it we take a standard greedy algorithm (similar to the matching pursuit from the compressed sensing literature) and turn it into a genetic algorithm by adding the local search and crossover steps. The main variable in this algorithm is the family (population) of the sets A which is denoted by $\mathcal{A}$. The cap on the population size is set to be $N_{\text{cap}} > 1$. The algorithm is as follows:

1. Let $\mathcal{A} = \{\emptyset\}$ as the solution for $N_{\text{nz}} = 0$.
2. For each $A \in \mathcal{A}$ find $i \notin A$, such that c corresponding to $\mathcal{A} \cup \{i\}$ is the best (i.e., minimizing $\|Xc - g\|$). Then replace A with $A \cup \{i\}$.
3. "Crossover": If $|\mathcal{A}| > 1$, then do the following. For each pair of sets, $A, A' \in \mathcal{A}$, divide randomly these sets into two nonintersecting subsets, $A = A_1 \cup A_2$ and $A' = A'_1 \cup A'_2$, and generate new sets $A_1 \cup A'_2$ and $A'_1 \cup A_2$. To generate such splittings of the sets, first sample uniformly an integer $m \in \{1, \dots, |A \setminus A'|\}$ and then form A_2 and A'_2 by uniformly sampling m distinct elements from A and A', respectively. Then replace all the sets in $\mathcal{A}$ with the newly generated sets. (Note that if $|\mathcal{A}| = N_{\text{cap}}$, then up to $N_{\text{cap}}(N_{\text{cap}} - 1)$ sets will be generated—two sets from each of $\binom{N_{\text{cap}}}{2}$ pairs.)
4. "Local search":
   4.1. For each $A \in \mathcal{A}$ find $j \in A$ such that c corresponding to $A \setminus \{j\}$ is the best.
   4.2. Then find $i \notin A \setminus \{j\}$ such that c corresponding to $(A \setminus \{j\}) \cup \{i\}$ is the best.
   4.3. If $i \neq j$, then
       4.3a. include $(A \setminus \{j\}) \cup \{i\}$, into $\mathcal{A}$
       4.3b. if $|\mathcal{A}| > N_{\text{cap}}$, then exclude A from $\mathcal{A}$, and
       4.3c. go to step 4.1.
5. Remove all but N_{cap} best sets in $\mathcal{A}$.
6. Repeat steps 2–5 until the accuracy goal on $\|Xc - g\|$ is reached. Then take the best set $A \in \mathcal{A}$ and compute the corresponding c.

We note that whenever A is fixed, then finding c corresponding to A is easy: $c = ((X^T X)_{AA})^{-1} (X^T g)_A$, where $\bullet_A$ and $\bullet_{AA}$ are the operations of extracting a subset of rows and columns corresponding to $A \subset \boldsymbol{A}$.

5. Numerical experiments. Our next goal is to understand how MTP performs compared to other interatomic potentials. Unfortunately, there are no existing works performing quantitative comparison between different machine learning potentials in terms of their accuracy and computational efficiency. In the present work we compare the performance of MTP with that of GAP for tungsten [26, 27] on the QM database published at www.libatoms.org together with the GAP code. We test MTP by fitting it on the database of 9693 configurations of tungsten, with nearly 150,000 individual atomic environments, and compare it to the analogously fitted GAP, namely, the one tagged as GAP_6 in [27] or as the iterative-SOAP-GAP or I-S-GAP in [26]. We note that this is a database of the Kohn–Sham DFT calculations (as opposed to a tight-binding model used in the analysis) for an electronic temperature of $1000°$K. We did not prove algebraic convergence of MTP to this model; however, we will observe it numerically.

We choose $R_{\text{cut}} = 4.9\text{Å}$ and also set the minimal distance parameter to be $R_{\min} :=$ 1.9Å. For the radial functions we choose

$$\hat{f}_{\mu,\nu}(r) := \begin{cases} r^{-\nu-2} r^{\mu} (R_{\text{cut}} - r)^2, & r < R_{\text{cut}}, \\ 0, & r \geq R_{\text{cut}}, \end{cases}$$

and then for each ν we orthonormalize them on the interval $[R_{\min}, R_{\text{cut}}]$ with the weight $r^{2\nu}(r - R_{\min})(R_{\text{cut}} - r)$. This procedure yields us the functions $f_{\mu,\nu}$ used with (4.1). (This procedure is equivalent to first orthonormalizing the functions $r^{-2} r^{\mu} (R_{\text{cut}} - r)^2$ with the weight $(r - R_{\min})(R_{\text{cut}} - r)$, the same weight as for the Chebyshev polynomials, and then multiplying by $r^{-\nu}$.) Here $r^{-\nu}$ compensates for $r^{\otimes\nu}$, r^{-2} prioritizes closer atoms to more distant onces, and $(R_{\text{cut}} - r)^2$ ensures a smooth cut-off.

5.1. Convergence with the number of basis functions. Even though theoretically it is proved that MTP is systematically improvable, the actual accuracy depends on the size of the regression problem to be solved. We therefore first study how fast the MTP fit converges with the number of basis functions used.

Theorem 3.2 suggests that the fitting error decreases exponentially with the polynomial degree. At the same time, a crude upper bound on the dimension of the space of polynomials of degree m is $(n+1)^m$, where n is the maximal number of atoms in an atomic neighborhood. This suggests an algebraic decay of the fitting error with the number of basis functions.

We hence study convergence of the error of fitting of potentials based on two choices of the sets of α. The first choice of the set of α, A_N, is to limit the corresponding degree of B_α for $\alpha \in A_N$ by N:

$$A_N = \{\alpha : \deg(B_\alpha) \leq N\}.$$

Since $f_{\mu,\nu}$ are no longer polynomials, we make a convention that the degree of $f_{\mu,\nu}$ is $\mu + 1$ (while the degree of the constant function $f(u) := 1$ is zero). Interestingly, it was found that a slightly better choice is

$$A'_N = \{\alpha : \deg(B_\alpha) \leq N + 8(\#\alpha)\},$$

where $\#\alpha$ is the length of α. In fact, A'_N corresponds to A_N if we make another convention that $\deg(f_{\mu,\nu}) = \mu + 5$.

Figure 5.1 displays the RMS fitting error in forces as a function of the size of the set A_N. One can observe an algebraic convergence, which indicates that Theorem 3.2 is valid for the Kohn–Sham DFT also. The observed rate is $(\#A)^{-0.227}$, where the exponent -0.227 is not a universal constant associated with MTP, but depends on the database chosen. A preasymptotic regime with a faster algebraic convergence rate can be seen for small $\#A$.

5.2. Performance tests. We next test the performance of MTP and compare it to GAP in terms of fitting error and computation (CPU) time. We choose two versions of MTP different from each other by the set A.

The first MTP, denoted by MTP_1, is generated by limiting $\deg(B_\alpha) + 8(\#\alpha) \leq 62$ and additionally limiting $(\#\alpha) \leq 4$, $\mu \leq 5$, and $\nu \leq 4$. MTP_2 is generated by limiting $\deg(B_\alpha) + 8(\#\alpha) \leq 52$, additionally limiting $(\#\alpha) \leq 5$, $\mu \leq 3$, and $\nu \leq 5$, and applying the ℓ_0 regularization algorithm with $N_{\text{cap}} = 4$, as described in section 4.2.2, to extract 760 basis functions (so as to make the accuracy of GAP and MTP_2 the same, as discussed in the next paragraph).
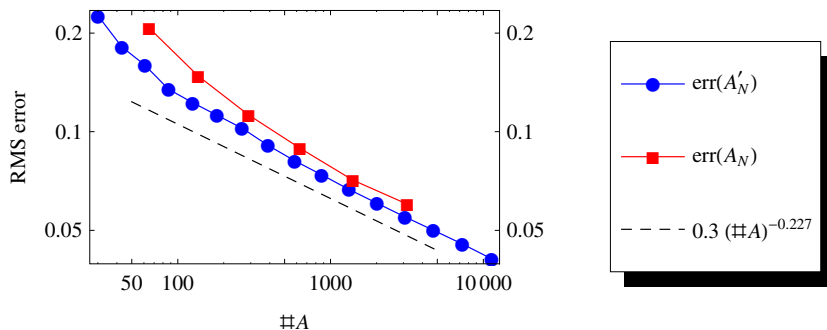
FIG. 5.1. *RMS fitting error in forces as a function of the size of A_N and A'_N. An algebraic convergence can be observed.*

TABLE 5.1
Efficiency and accuracy of MTP as compared to GAP. The mean CPU time is given for a single-core computation on an Intel i7-2675QM laptop CPU, together with the standard deviation as computed on a number of independent runs. The last section of the table reports the force RMS errors for the 16-fold cross-validation.

Potential	GAP	MTP_1	MTP_2
CPU time/atom [ms]	134.2 $\pm$2.6	2.9 $\pm$0.5	0.8 $\pm$0.2
basis functions	10 000	11 133	760
Fit errors			
force RMS error [eV/Å]	0.0633	0.0427	0.0633
[%]	4.2%	2.8%	4.2%
Cross-validation errors			
regularization parameter γ	-	$3 \cdot 10^{-9}$	0
force RMS error[eV/Å]	-	0.0511	0.0642
[%]	-	3.4%	4.3%

The data from the conducted efficiency and accuracy tests are summarized in Table 5.1. The RMS force (more precisely, RMS of all force components over all non-single-atom configurations) is 1.505 eV/Å, which is used to compute the relative RMS error. The errors relative to this RMS force are also presented in the table. The GAP error is calculated based on the data from [26]. The CPU times do not include the initialization (precomputation) or constructing the atomistic neighborhoods.

One can see that MTP_1 has about the same number of fitting parameters as GAP, while its fitting accuracy is about 1.5 times better and the computation time is 40 times smaller. MTP_2 was constructed such that its fitting accuracy is the same as GAP, but it uses many fewer parameters for fitting and its computation is more than two orders of magnitude faster.

Also included in the table is the 16-fold cross-validation error. It shows that MTP_2 is not overfitted on the given database, whereas MTP_1 needs regularization to avoid overfitting. We anticipate, however, that by significantly increasing the database size, the cross-validation error of MTP_1 would go down and reach the current fitting error of MTP_1, since the fitting error follows closely the algebraic decay (see Figure 5.1) and is not expected to deteriorate with increasing the database size.

6. Conclusion. The present paper considers the problem of approximating a QM interaction model with interatomic potentials from a mathematical point of view. In particular, (1) a new class of nonparametric (i.e., systematically improvable) po-

tentials satisfying all the required symmetries has been proposed and advantages in terms of accuracy and performance over the existing schemes have been discussed and (2) an algebraic convergence of fitting with these potentials in a simple setting has been proved and it was then confirmed with the numerical experiments.

This work is done under the assumption that all atoms are chemically equivalent. A straightforward extension to multicomponent systems would be to let the radial functions depend not only on the positions of atoms x_i but also on the types of atoms t_i Thus, the expression for the moments for atom i could be

$$M_{\mu,\nu} = \sum_j f_{\mu,\nu}(|x_j - x_i|, t_i, t_j)(x_j - x_i)^{\otimes\nu},$$

where the summation is over the neighborhood of atom i. We leave exploring this path to future publications.

REFERENCES

[1] X. Ai, Y. Chen, and C. A. Marianetti, *Slave mode expansion for obtaining ab initio interatomic potentials*, Phys. Rev. B, 90 (2014), 014308.

[2] E. Alpaydin, *Introduction to Machine Learning*, MIT Press, Cambridge, MA, 2014.

[3] A. P. Bartók, R. Kondor, and G. Csányi, *On representing chemical environments*, Phys. Rev. B, 87 (2013), 184115.

[4] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, *Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons*, Phys. Rev. Lett., 104 (2010), 136403.

[5] F. L. Bauer and C. T. Fike, *Norms and exclusion theorems*, Numer. Math., 2 (1960), pp. 137–141.

[6] J. Behler, *Atom-centered symmetry functions for constructing high-dimensional neural network potentials*, J. Chemical Phys., 134 (2011), 074106.

[7] J. Behler, *Representing potential energy surfaces by high-dimensional neural network potentials*, J. Phys. Condensed Matter, 26 (2014), 183001.

[8] J. Behler and M. Parrinello, *Generalized neural-network representation of high-dimensional potential-energy surfaces*, Phys. Rev. Lett., 98 (2007), 146401.

[9] B. J. Braams and J. M. Bowman, *Permutationally invariant potential energy surfaces in high dimensionality*, Internat. Rev. Phys. Chemistry, 28 (2009), pp. 577–606.

[10] E. Cances, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday, *Computational quantum chemistry: A primer*, in Computational Chemistry, Handb. Numer. Anal., 10, P. G. Ciarlet and C. L. Bris, Eds., Elsevier, Amsterdam, 2003, pp. 3–270.

[11] H. Chen and C. Ortner, *QM/MM methods for crystalline defects. Part 1: Locality of the tight binding model*, Multiscale Model. Simul., 14 (2016), pp. 232–264.

[12] H. Chen and C. Ortner, *QM/MM Methods for Crystalline Defects. Part 2: Consistent Energy and Force-Mixing*, preprint, arXiv:1509.06627, 2015.

[13] G. Ferré, J.-B. Maillet, and G. Stoltz, *Permutation-invariant distance between atomic configurations*, J. Chemical Phys., 143 (2015), 104114.

[14] M. Finnis, *Interatomic Forces in Condensed Matter*, Vol. 1, Oxford University Press, Oxford, UK, 2003.

[15] M. Griebel, S. Knapek, and G. Zumbusch, *Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications*, Vol. 5, Springer Science & Business Media, New York, 2007.

[16] W. Hackbusch and B. N. Khoromskij, *Towards $\mathcal{H}$-matrix approximation of the linear complexity*, in Problems and Methods in Mathematical Physics, Oper. Theory Adv. Appl. 121, Birkhäuser Verlag, Basel, 2001, pp. 194–220.

[17] O. Hellman, P. Steneteg, I. A. Abrikosov, and S. I. Simak, *Temperature dependent effective potential method for accurate free energy calculations of solids*, Phys. Rev. B, 87 (2013), 104111.

[18] N. J. Higham, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, 2008.

[19] M. Hirn, N. Poilvert, and S. Mallat, *Quantum Energy Regression Using Scattering Transforms*, preprint, arXiv:1502.02077, 2015.

[20] J. Jalkanen and M. H. Müser, *Systematic analysis and modification of embedded-atom potentials: Case study of copper*, Model. Simul. Materials Sci. Eng., 23 (2015), 074001.

[21] Z. Li, J. R. Kermode, and A. De Vita, *Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces*, Phys. Rev. lett., 114 (2015), 096405.

[22] S. Linnainmaa, *Taylor expansion of the accumulated rounding error*, BIT, 16 (1976), pp. 146–160.

[23] F. Q. Nazar and C. Ortner, *Locality of the Thomas-Fermi-von Weizsäcker Equations*, preprint, arXiv:1509.06753, 2015.

[24] L. J. Nelson, V. Ozoliņš, C. S. Reese, F. Zhou, and G. L. W. Hart, *Cluster expansion made easy with Bayesian compressive sensing*, Phys. Rev. B, 88 (2013), 155105.

[25] G. Szegö, *Orthogonal Polynomials*, Vol. 23, AMS, Providence, RI, 1939.

[26] W. J. Szlachta, *First Principles Interatomic Potential for Tungsten Based on Gaussian Process Regression*, preprint, arXiv:1403.3291, 2014.

[27] W. J. Szlachta, A. P. Bartók, and G. Csányi, *Accuracy and transferability of Gaussian approximation potential models for tungsten*, Phys. Rev. B, 90 (2014), 104108.

[28] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, *SNAP: Automated generation of quantum-accurate interatomic potentials*, J. Comput. Phys., 285 (2015), pp. 316–330.

[29] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B Methodol., 58 (1996), pp. 267–288.

[30] H. Weyl, *The Classical Groups: Their Invariants and Representations*, Princeton Landmarks in Mathematics and Physics, Princeton University Press, Princeton, NJ, 1997.

[31] Y. Zhang, *Theory of compressive sensing via ℓ_1-minimization: A non-rip analysis and extensions*, J. Oper. Res. Soc. China, 1 (2013), pp. 79–105.