Accurate interatomic force fields via machine learning with covariant kernels

Aldo Glielmo,[1,*] Peter Sollich,[2] and Alessandro De Vita[1,3]
[1]*Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom*
[2]*Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom*
[3]*Dipartimento di Ingegneria e Architettura, Università di Trieste, via A. Valerio 2, I-34127 Trieste, Italy*
(Received 4 November 2016; revised manuscript received 13 April 2017; published 8 June 2017)

We present a novel scheme to accurately predict atomic forces as vector quantities, rather than sets of scalar components, by Gaussian process (GP) regression. This is based on matrix-valued kernel functions, on which we impose the requirements that the predicted force rotates with the target configuration and is independent of any rotations applied to the configuration database entries. We show that such covariant GP kernels can be obtained by integration over the elements of the rotation group $SO(d)$ for the relevant dimensionality d. Remarkably, in specific cases the integration can be carried out analytically and yields a conservative force field that can be recast into a pair interaction form. Finally, we show that restricting the integration to a summation over the elements of a finite point group relevant to the target system is sufficient to recover an accurate GP. The accuracy of our kernels in predicting quantum-mechanical forces in real materials is investigated by tests on pure and defective Ni, Fe, and Si crystalline systems.

DOI: 10.1103/PhysRevB.95.214302

I. INTRODUCTION

Recent decades have witnessed an exponential growth of computer processing power (Moore's law [1]) and an equally fast progress in storage technology (Kryder's law [2,3]). Atomistic modeling methods based on computation and data-intensive quantum-mechanical methods, such as density functional theory (DFT) [4–6], have correspondingly evolved in both feasibility and scope. Moreover, the possibility of retaining at low cost very large amounts of data generated by quantum-mechanical (QM) codes has prompted novel efforts to make the data openly accessible [7].

The information contained in the data can thus be harnessed and reused indefinitely, in various ways. High-throughput techniques are routinely used to identify new correlations between physical properties, with the aim of designing new high-performance materials [8–10]. Inference techniques can meanwhile also be used as a boost or substitute for QM techniques. This typically involves predicting a physical property for a new system configuration, on the basis of its values for an existing database of configurations. If the database is sufficiently large and representative, the new property values can be quickly inferred, rather than calculated anew by expensive QM procedures, with controllable accuracy.

Machine learning techniques have been successfully used to predict properties as diverse as atomization energies [11], density functionals [12], Green's functions [13], electronic transport coefficients [14], potential energy surfaces [15–17], and free-energy landscapes [18]. The high configuration space complexity of real chemical systems has also inspired learning molecular dynamics schemes that never assume database completeness, but rather combine inference with on-the-fly QM calculations [learning on the fly (LOTF)] [19–21] carried out when inference is infeasible or not deemed sufficiently accurate.

A well-established general concept within the machine learning community is that functional invariance properties under some known transformation can be used to improve prediction, whether this is carried out by, e.g., Gaussian process (GP) regression [22,23] or neural networks [24]. Exploiting in similar ways properties other than invariance has received more limited attention [25]. In the same spirit, materials modelers have been successful in exploiting the invariance of energy under rotation or translation to improve the performance of energy prediction techniques [15,16]. In LOTF molecular dynamics applications the high-accuracy target and local interpolation character of force prediction makes it appealing to learn forces directly rather than learning a potential energy scalar field first and then deriving forces by differentiation. In previous works [26–28] this was accomplished by using GP regression to separately learn individual force components.

Here, we show how vectorial Gaussian process (VGP) [29,30] regression provides a more natural framework for force learning, where the correct vector behavior of forces under symmetry transformations can be obtained by using a new family of vector kernels of covariant nature. These kernels prove particularly efficient at exploiting the information contained in QM force databases, however constructed, together with any prior knowledge of the symmetry properties of the physical system under investigation. The next section provides a brief overview of the notion of a VGP, where we pay particular attention to the problem of force learning. Then we define a covariant kernel, explain its symmetry properties, and give a general recipe to generate such kernels. The procedure is best exemplified by looking at one- and two-dimensional (2D) systems, where the relevant symmetry force transformation groups are D_1 and $O(2)$. Finally, we address the full three-dimensional case, where covariant kernels are tested by examining their performance in learning QM forces in realistic physical systems [31].

II. VECTORIAL GAUSSIAN PROCESS REGRESSION

We wish to model by a VGP the force $\mathbf{f}$ acting on an atom whose chemical environment is in a configuration ρ that encodes the positions of all neighbors of the atom, up to a suitable

*aldo.glielmo@kcl.ac.uk

cutoff radius, in an arbitrary Cartesian reference frame. In the absence of long-range ionic interactions, the existence of such a local map is guaranteed for all finite-temperature systems by the nearsightedness principle of electronic matter [32,33].

In a Bayesian setting, before any data is considered, $\mathbf{f}$ is treated as a Gaussian process, i.e., it is assumed that for any finite set of configurations $\{\rho_i, i = 1, \dots N\}$ the values $\mathbf{f}(\rho_i)$ taken by the vector function $\mathbf{f}$ are well described by a multivariate Gaussian distribution [23]. We write:

$$\mathbf{f}(\rho) \sim \mathcal{GP}(\mathbf{m}(\rho), \mathbf{K}(\rho, \rho')), \tag{1}$$

where $\mathbf{m}(\rho)$ is a vector-valued mean function and $\mathbf{K}(\rho, \rho')$ is a matrix-valued kernel function. Before any data is considered, $\mathbf{m}$ is usually assumed to be zero as all prior information on $\mathbf{f}$ is encoded into the kernel function $\mathbf{K}(\rho, \rho')$. The latter represents the correlation of the vectors $\mathbf{f}(\rho)$ and $\mathbf{f}(\rho')$ as a function of the two configurations (input space points) ρ and ρ':

$$\mathbf{K}(\rho, \rho') = \langle \mathbf{f}(\rho) \mathbf{f}^{\mathrm{T}}(\rho') \rangle, \tag{2}$$

where angular brackets here signify the expected value over the multivariate Gaussian distribution. Any kernel $\mathbf{K}$ consistent with this definition must be a positive semidefinite matrix function, since for any collection of vectors $\{\mathbf{v}_i\}$

$$\sum_{ij} \mathbf{v}_i^{\mathrm{T}} \mathbf{K}(\rho_i, \rho_j) \mathbf{v}_j = \left\langle \left(\sum_i \mathbf{v}_i^{\mathrm{T}} \mathbf{f}(\rho_i) \right)^2 \right\rangle \geqslant 0. \tag{3}$$

To train the prediction model we need to access a database of atomic configurations and reference forces $\mathcal{D} = \{(\rho, \mathbf{f}^r)_i, i = 1, \dots, N\}$. Using Bayes' theorem [34] the distribution (1) is modified to take the data $\mathcal{D}$ into account [23]. If the likelihood function [24] is also Gaussian (which effectively assumes that the observed forces $\mathbf{f}_i^r$ are the true forces subject to Gaussian noise of variance σ_n^2) then the resulting posterior distribution $\mathbf{f}(\rho|\mathcal{D})$, conditional on the data, will also be a Gaussian process

$$\mathbf{f}(\rho|\mathcal{D}) \sim \mathcal{GP}(\hat{\mathbf{f}}(\rho|\mathcal{D}), \hat{\mathbf{C}}(\rho, \rho')). \tag{4}$$

The mean function of the posterior distribution, $\hat{\mathbf{f}}(\rho|\mathcal{D})$, is at this point the best estimate for the true underlying function:

$$\hat{\mathbf{f}}(\rho|\mathcal{D}) = \sum_{ij}^{N} \mathbf{K}(\rho, \rho_i) \left[\mathbb{K} + \mathbb{I}\sigma_n^2 \right]_{ij}^{-1} \mathbf{f}_j^r. \tag{5}$$

Here σ_n^2, formally the noise affecting the observed forces $\mathbf{f}^r$, serves in practice as a regularizer for the matrix inverse. In the following, blackboard bold characters such as $\mathbb{K}$ or $\mathbb{I}$ indicate $N \times N$ block matrices [for instance, the Gram matrix $\mathbb{K}$ is defined as $(\mathbb{K})_{ij} = \mathbf{K}(\rho_i, \rho_j)$]. Similarly, we denote by $[\mathbb{K} + \mathbb{I}\sigma_n^2]_{ij}^{-1}$ the ij block of the inverse matrix.

We next examine how to incorporate the vector behavior of forces into the learning algorithm. The relevant symmetry transformations in the input space are: rigid translation of all atoms, permutation of atoms of the same chemical species, rotations and reflections of atomic configurations. Forces are invariant with respect to translations and atomic permutations, and covariant with respect to rotations and reflections. Assuming that the representation of the atomic configuration is local, i.e., the atom subject to the force $\mathbf{f}_i$ is at the origin of the reference frame used for ρ_i, translations are automatically taken into account. The remaining symmetries must be addressed in the construction of covariant kernels.

III. COVARIANT KERNELS

From now on we will define $\mathcal{S}$ to be any symmetry operator (rotation or reflection) acting on an atomistic configuration of a d-dimensional system. Rotations will be denoted by $\mathcal{R}$ and reflections by $\mathcal{Q}$.

We require two properties to apply to the predicted force $\hat{\mathbf{f}}(\rho|\mathcal{D})$, once configurations are transformed by an operator $\mathcal{S}$ (represented by a matrix $\mathbf{S}$):

Property 1. If the target configuration ρ is transformed to $\mathcal{S}\rho$, the predicted force must transform accordingly:

$$\hat{\mathbf{f}}(\mathcal{S}\rho|\mathcal{D}) = \mathbf{S}\hat{\mathbf{f}}(\rho|\mathcal{D}). \tag{6}$$

Property 2. The predicted force must not change if we arbitrarily transform the configurations in the database ($\mathcal{D} \to \tilde{\mathcal{D}} = \{(\mathcal{S}_i\rho_i, \mathbf{S}_i\mathbf{f}_i^r)\}$) with any chosen set of rotoreflections $\{\mathcal{S}_i\}$.

We next introduce a special class of kernel functions that automatically guarantees these two properties: a covariant kernel has the defining property

$$\mathbf{K}(\mathcal{S}\rho, \mathcal{S}'\rho') = \mathbf{S}\mathbf{K}(\rho, \rho')\mathbf{S}'^{\mathrm{T}}. \tag{7}$$

That a covariant kernel imposes Property 1 follows straightforwardly from Eq. (5):

$$\begin{aligned} \hat{\mathbf{f}}(\mathcal{S}\rho|\mathcal{D}) &= \sum_{ij}^{N} \mathbf{K}(\mathcal{S}\rho, \rho_i) \left[\mathbb{K} + \mathbb{I}\sigma_n^2 \right]_{ij}^{-1} \mathbf{f}_j^r \\ &= \sum_{ij}^{N} \mathbf{S}\mathbf{K}(\rho, \rho_i) \left[\mathbb{K} + \mathbb{I}\sigma_n^2 \right]_{ij}^{-1} \mathbf{f}_j^r \\ &= \mathbf{S}\hat{\mathbf{f}}(\rho|\mathcal{D}). \end{aligned} \tag{8}$$

To prove Property 2 we note that, if the kernel function is covariant, the transformed database $\tilde{\mathcal{D}}$ has Gram matrix $(\tilde{\mathbb{K}})_{ij} = \mathbf{K}(\mathcal{S}_i\rho_i, \mathcal{S}_j\rho_j) = \mathbf{S}_i\mathbf{K}(\rho_i, \rho_j)\mathbf{S}_j^{\mathrm{T}}$. If we define the block-diagonal matrix $\mathbb{S}_{ij} = \delta_{ij}\mathbf{S}_i$, this can be written in the simple block-matrix form $\tilde{\mathbb{K}} = \mathbb{S}\mathbb{K}\mathbb{S}^{\mathrm{T}}$. Using kernel covariance again to write $\mathbf{K}(\rho, \mathcal{S}_i\rho_i) = \mathbf{K}(\rho, \rho_i)\mathbb{S}_{ii}^{\mathrm{T}}$ the prediction associated with the transformed database $\tilde{\mathcal{D}}$ can be written as

$$\hat{\mathbf{f}}(\rho|\tilde{\mathcal{D}}) = \sum_{ij}^{N} \mathbf{K}(\rho, \rho_i)\mathbb{S}_{ii}^{\mathrm{T}} \left[\mathbb{S}\mathbb{K}\mathbb{S}^{\mathrm{T}} + \mathbb{I}\sigma_n^2 \right]_{ij}^{-1} \mathbb{S}_{jj}\mathbf{f}_j^r. \tag{9}$$

By simple matrix manipulations it is now possible to show that in the above expression the symmetry transformations cancel out; indeed

$$\begin{aligned} \mathbb{S}^{\mathrm{T}} \left[\mathbb{S}\mathbb{K}\mathbb{S}^{\mathrm{T}} + \mathbb{I}\sigma_n^2 \right]^{-1} \mathbb{S} &= \mathbb{S}^{\mathrm{T}} \left[\mathbb{S}\left(\mathbb{K} + \mathbb{I}\sigma_n^2 \right) \mathbb{S}^{\mathrm{T}} \right]^{-1} \mathbb{S} \\ &= \mathbb{S}^{\mathrm{T}} (\mathbb{S}^{\mathrm{T}})^{-1} \left[\mathbb{K} + \mathbb{I}\sigma_n^2 \right]^{-1} \mathbb{S}^{-1} \mathbb{S} \\ &= \left[\mathbb{K} + \mathbb{I}\sigma_n^2 \right]^{-1}. \end{aligned} \tag{10}$$

Equation (10) along with (9) implies $\hat{\mathbf{f}}(\rho|\tilde{\mathcal{D}}) = \hat{\mathbf{f}}(\rho|\mathcal{D})$, that is, Property 2. It is easy to check that standard kernels such as the squared exponential [24] or the overlap integral of atomic configuration [35] do not possess the covariance property (7). Designing, entirely by feature engineering, a covariant kernel

is in principle possible but can require complex tuning and is likely to be highly system dependent (see, e.g., Ref. [26]). We note that noncovariant kernels can be used and avoid these difficulties, and some have been successfully implemented [28,36]. This leaves space for improvement as prediction efficiency will generally be enhanced by increased exploitation of symmetry (see, e.g., Fig. 3 below for a simple test of this).

We next present a general method for transforming a standard matrix kernel into a covariant one, followed by numerical tests suggesting that the resulting kernel improves very significantly on the force-learning properties of the initial one, its error converging with just a fraction of the training data. This proceeds along the lines of previous techniques for generating scalar invariants, namely the transformation integration procedure developed in Ref. [22] and the smooth overlap of atomic orbitals (SOAP) representation for learning potential energy surfaces of atomic systems [37,38].

Given a group $\mathcal{S}$ and a base kernel $\mathbf{K}^b$, a covariant kernel $\mathbf{K}^c$ can be constructed by

$$\mathbf{K}^c(\rho,\rho') = \int d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}} \mathbf{K}^b(\mathcal{S}_1\rho,\mathcal{S}_2\rho')\mathbf{S}_2, \tag{11}$$

where $d\mathcal{S}$ is the normalized Haar measure for the symmetry group we are integrating over [39].

The covariance of $\mathbf{K}^c$ as given by (11) is easily checked as

$$\begin{aligned}\mathbf{K}^c(\mathcal{S}\rho,\mathcal{S}'\rho') &= \int d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}} \mathbf{K}^b(\mathcal{S}_1\mathcal{S}\rho,\mathcal{S}_2\mathcal{S}'\rho')\mathbf{S}_2 \\ &= \int d\tilde{\mathcal{S}}_1 d\tilde{\mathcal{S}}_2\, \mathbf{S}\tilde{\mathbf{S}}_1^{\mathrm{T}} \mathbf{K}^b(\tilde{\mathcal{S}}_1\rho,\tilde{\mathcal{S}}_2\rho')\tilde{\mathbf{S}}_2\mathbf{S}'^{\mathrm{T}} \\ &= \mathbf{S}\mathbf{K}^c(\rho,\rho')\mathbf{S}'^{\mathrm{T}},\end{aligned} \tag{12}$$

where the second line follows from the substitutions $\tilde{\mathcal{S}}_1 = \mathcal{S}_1\mathcal{S}$ and $\tilde{\mathcal{S}}_2 = \mathcal{S}_2\mathcal{S}'$. Note that these transformations have unit Jacobian because of the translational invariance (within the group) of any Haar measure [39,40].

It can be shown that the positive semidefiniteness of the base kernel is preserved under the operation (11) of covariant integration. In particular, a kernel is positive semidefinite if and only if it is a scalar product in some (possibly infinite dimensional) vector space [23,41]. Hence the base kernel can be written as $\mathbf{K}^b(\rho,\rho') = \int d\alpha\, \boldsymbol{\phi}_\alpha(\rho)\boldsymbol{\phi}_\alpha^{\mathrm{T}}(\rho')$. It is then possible to show that its covariant counterpart $\mathbf{K}^c$ [Eq. (11)] will also be a scalar product in a new function space. Indeed

$$\begin{aligned}\mathbf{K}^c(\rho,\rho') &= \int d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}} \mathbf{K}^b(\mathcal{S}_1\rho,\mathcal{S}_2\rho')\mathbf{S}_2 \\ &= \int d\alpha\, d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}} \boldsymbol{\phi}_\alpha(\mathcal{S}_1\rho)\boldsymbol{\phi}_\alpha^{\mathrm{T}}(\mathcal{S}_2\rho')\mathbf{S}_2 \\ &= \int d\alpha\, \boldsymbol{\psi}_\alpha(\rho)\boldsymbol{\psi}_\alpha^{\mathrm{T}}(\rho'),\end{aligned} \tag{13}$$

where the new basis vectors were defined as $\boldsymbol{\psi}_\alpha(\rho) = \int d\mathcal{S}\, \mathbf{S}^{\mathrm{T}}\boldsymbol{\phi}_\alpha(\mathcal{S}\rho)$. Hence, $\mathbf{K}^c$ will also be positive definite.

The completely general procedure above can be cumbersome to apply in practice, because of the double integration over group elements in (11) and the dependence on the design of the base kernel matrix $\mathbf{K}^b$. As a simplification, we assume the base kernel to be of diagonal form; assuming equivalence of all space directions, we can then write

$$\mathbf{K}^b(\rho,\rho') = \mathbf{I}k^b(\rho,\rho'), \tag{14}$$

where the scalar base kernel k^b is independent on the reference frame in which the configurations are expressed. This requires that

$$k^b(\mathcal{S}\rho,\mathcal{S}\rho') = k^b(\rho,\rho'), \tag{15}$$

that is, scalar invariance of the base kernel (a property very commonly found in standard kernels). The double integration in (11) reduces at this point to a single one

$$\begin{aligned}\mathbf{K}^c(\rho,\rho') &= \int d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}}\mathbf{S}_2 k^b(\mathcal{S}_1\rho,\mathcal{S}_2\rho') \\ &= \int d\mathcal{S}_1 d\mathcal{S}_2\, \mathbf{S}_1^{\mathrm{T}}\mathbf{S}_2 k^b\left(\rho,\mathcal{S}_1^{-1}\mathcal{S}_2\rho'\right) \\ &= \int d\mathcal{S}\, \mathbf{S}\, k^b(\rho,\mathcal{S}\rho'),\end{aligned} \tag{16}$$

where the second line follows from property (15) and the third line is obtained by the substitution $\mathcal{S} = \mathcal{S}_1^{-1}\mathcal{S}_2$.

In the next section we show that some base kernels allow analytical integration of (16). Here we note that incorporating our prior knowledge of the correct behavior of forces in the kernel enables us to learn and predict forces associated with any configuration, regardless of its orientation. However, being able to do this for completely generic orientations is not always necessary. In many systems (e.g., crystalline solids where the orientation is known) all relevant configurations cluster around particular discrete symmetries. For these systems the relevant physics can be captured by restricting Eq. (11) to a discrete sum over the relevant group elements:

$$\mathbf{K}^c(\rho,\rho') = \frac{1}{|G|}\sum_{\mathcal{G}\in G} \mathbf{G}k(\rho,\mathcal{G}\rho'), \tag{17}$$

and since there are 48 distinct group elements at most (the order of the full O_{48} group), the procedure remains computationally feasible. In the particular case of one-dimensional systems, where the only available symmetry operation other than the identity is the inversion, Eqs. (16) and (17) are formally equivalent.

IV. COVARIANT KERNELS FROM ONE TO THREE DIMENSIONS

In the following we will assume that a single chemical species is present, so that permutation invariance will be simply enforced by representing configurations as linear combinations of n Gaussian functions each centered on one atom, all having the same width σ, and suitably normalized depending on the dimension d considered:

$$\rho(\mathbf{r},\{\mathbf{r}_i\}) = \frac{1}{(2\pi\sigma^2)^{d/2}}\sum_i^n e^{-\frac{\|\mathbf{r}-\mathbf{r}_i\|^2}{2\sigma^2}}. \tag{18}$$

From (18), a linear base kernel k_L^b can be defined as the overlap integral of two configurations [15,35]

$$\begin{aligned} k_L^b(\rho,\rho') &= \int d\mathbf{r}\, \rho(\mathbf{r},\{\mathbf{r}_i\})\rho'(\mathbf{r},\{\mathbf{r}'_j\}) \\ &= \frac{1}{(2\pi\sigma^2)^d} \sum_{ij}^{nn'} \int d\mathbf{r}\, e^{-\frac{\|\mathbf{r}-\mathbf{r}_i\|^2}{2\sigma^2}} e^{-\frac{\|\mathbf{r}-\mathbf{r}'_j\|^2}{2\sigma^2}} \\ &= \frac{1}{(2\sqrt{\pi\sigma^2})^d} \sum_{ij} e^{-\frac{\|\mathbf{r}_i-\mathbf{r}'_j\|^2}{4\sigma^2}}, \end{aligned} \tag{19}$$

where the integration yielding the third line is performed by standard completion of the square.

We can interpret the linear kernel k_L^b in (19) as a scalar product in function space, so that $k_L^b(\rho,\rho) = \|\rho\|^2$ can be thought of as the squared norm of the ρ configuration function. A permutation invariant distance is also readily obtained as $d(\rho,\rho') = \|\rho - \rho'\|$, which can be used within a squared exponential kernel to give

$$\begin{aligned} k_{SE}^b(\rho,\rho') &= e^{-\|\rho-\rho'\|^2/2\theta} \\ &= e^{-(k_L^b(\rho,\rho)+k_L^b(\rho',\rho')-2k_L^b(\rho,\rho'))/2\theta}. \end{aligned} \tag{20}$$

The representation described above is by construction translation (and atomic permutation) invariant. We next address the transformations for which the atomic force is covariant, i.e., rotations and reflections, using the approach described in the previous section. Systems with dimensions $d = 1,2,3$ are considered in the following three subsections. The first two provide a useful conceptual playground where the features of covariant learning can be more easily visualized. The third one benchmarks the method in real physical systems, simulated at the DFT level of accuracy.

A. 1D systems

A key feature of covariant kernels is the ability to enable learning of the entire set of configurations that are equivalent by symmetry to those actually provided in the database. For instance, the force acting on the (central) atom at the origin of configuration ρ can be predicted even if only configurations ρ' of different symmetry are contained in the database. The only relevant symmetry transformation in one dimension is the reflection $\mathcal{Q}$ of a configuration about its center. In the simplest possible system, a dimer, this maps configurations where the central atom has a right neighbor (i.e., those for which the central atom is the left atom in the dimer) onto configurations where the central atom has a left neighbor. The covariant symmetrisation discussed in the previous section [Eq. (17)] takes the very simple form

$$k^c(\rho,\rho') = \tfrac{1}{2}\big[k_L^b(\rho,\rho') - k_L^b(\rho,\mathcal{Q}\rho')\big]. \tag{21}$$

Note that k^c is identically zero for inversion-symmetric configurations ρ or ρ' whose associated forces must vanish.

The force field associated with a 1D Lennard Jones dimer is plotted in Fig. 1 (dotted curve) as a function of a single signed number—the 1D vector going from the central atom to its neighbor. The figure also shows the predictions of the unsymmetrized base kernel using training data coming from configurations centered on the left atom only (solid blue curve).

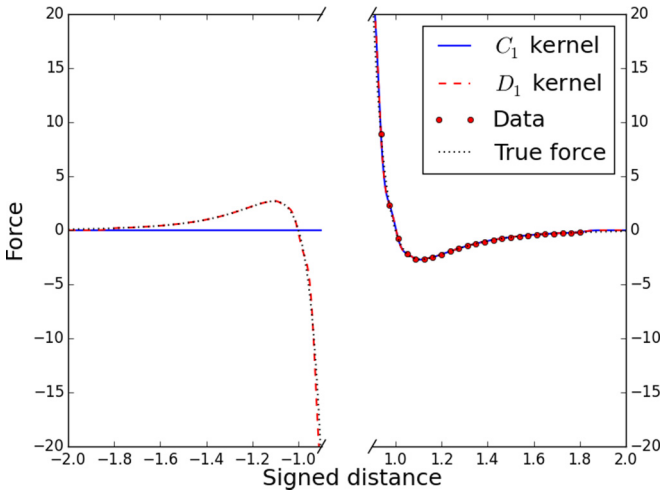

FIG. 1. Lennard-Jones dimer force field, learned with data from one atom only. The base kernel (C_1) does not learn the symmetric counterpart (reaction force), while the covariant (D_1) does. The kernels are labeled by the symmetry group used to make them covariant; see main text for details.

This closely reproduces the true LJ forces in the region where the data are available, and predicts the pure prior mean (i.e., zero) in the symmetry-related region, i.e., the left half of the figure. Meanwhile, because of the covariant constraint (prior information) the GP based on the covariant kernel learns the left part of the field by just reflecting the right part appropriately.

To further check the performance of the covariant kernel (21) we extended the comparison above to predicting the forces associated with a 1D Lennard Jones 50-atom chain system, in periodic boundary conditions. A database of training configurations and an independent test set of local configurations and forces were sampled from a constant temperature molecular dynamics simulation using a Langevin thermostat.

Before presenting the results, it is necessary to introduce some conventions that will apply throughout the rest of this work. As a measure of error between reference force $\mathbf{f}^r(\rho)$ and predicted force $\hat{\mathbf{f}}(\rho)$, we will take the absolute value of their vector difference $|\Delta\mathbf{f}| = |\mathbf{f}^r(\rho) - \hat{\mathbf{f}}(\rho)|$. Relative errors are obtained by dividing this absolute error by the time-ensemble average of the force modulus $|\bar{\mathbf{f}}|$. Average errors are found by randomly sampling N training configurations and 1000 test configurations. Repeating this operation provides the standard deviation and hence the error bars on absolute and relative errors. We furthermore denote by C_n the cyclic group of order n and by D_n the dihedral group (containing also reflections) of order $2n$ (C_1 hence indicates the trivial group).

With the above clarifications, we can proceed with the analysis of Fig. 2, which reports the average relative force error made by the GP regression on the test set as a function of training set size. It is immediately apparent that the covariant kernel performance is comparable to that of the base kernel with double the number of data points for training. We will observe the same effect also in two and three dimensions: symmetrizing over a relevant finite group of order $|G|$ gives rise

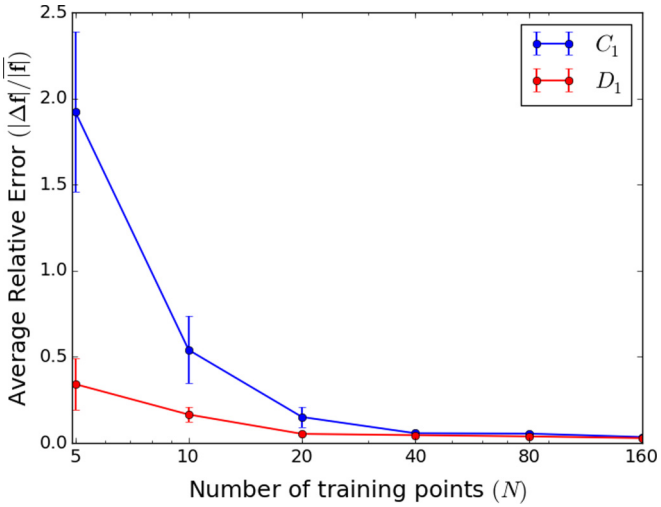

FIG. 2. Learning curves for a 1D chain of LJ atoms. The covariant kernel (D_1) learns twice as fast as the base one (C_1).

to an error drop approximately equivalent to a $|G|$-fold increase in the number of training points. Since the computational complexity of training a GP is $O(N^3)$, this can obviously lead to significant computer time savings.

B. 2D systems

In two dimensions all rotations and reflections, as well as any combination of these, are elements of $O(2)$. Moreover, the $O(2)$ group can be represented by the following set of matrices $O(2) = \{\mathbf{R}(\theta), \theta \in (0,2\pi]\} \cup \{\mathbf{R}(\theta)\mathbf{Q}, \theta \in (0,2\pi]\}$ where $\mathbf{R}(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$ and $\mathbf{Q}$ is any 2×2 reflection matrix.

This makes the covariant integration (16) over $O(2)$ trivial once the matrix elements resulting from the integration over $SO(2)$ have been calculated. We next carry out the integration for the linear base kernel of Eq. (19). This can be expressed as a sum of pair contributions, where the first atom in each pair belongs to ρ and the second to ρ' :

$$\mathbf{K}^c_{SO(2)}(\rho,\rho') = \frac{1}{L}\sum_{ij}^{nn'} \int_{SO(2)} d\mathcal{R}\, \mathbf{R}\, e^{-\frac{\|\mathbf{r}_i - \mathbf{R}\mathbf{r}'_j\|^2}{4\sigma^2}}. \quad (22)$$

Consistent with Eq. (16), only one atom of the pair is rotated during the integration, with L being the normalization factor [cf. Eq. (19)]. The pairwise integrals in (22) are calculated in two steps. We first define $\mathbf{R}_{ij}$ to be the rotation matrix which aligns $\mathbf{r}'_j$ onto $\mathbf{r}_i$, and then perform the change of variable $\tilde{\mathbf{R}} = \mathbf{R}\mathbf{R}^{\mathrm{T}}_{ij}$ (and analogously $\tilde{\mathcal{R}} = \mathcal{R}\mathcal{R}^{-1}_{ij}$) yielding

$$\mathbf{K}^c_{SO(2)}(\rho,\rho') = \frac{1}{L}\sum_{ij}\left(\int_{SO(2)} d\tilde{\mathcal{R}}\, \tilde{\mathbf{R}}\, e^{-\frac{\|\mathbf{r}_i - \tilde{\mathbf{R}}\mathbf{R}_{ij}\mathbf{r}'_j\|^2}{4\sigma^2}}\right)\mathbf{R}_{ij}. \quad (23)$$

Since the two vectors $\mathbf{r}_i$ and $\mathbf{R}_{ij}\mathbf{r}'_j$ are now aligned, the integral in Eq. (23) can only depend on the two moduli r_i and r'_j. The final result takes a very simple analytic form (cf. Supplemental Material [42]):

$$\mathbf{K}^c_{SO(2)}(\rho,\rho') = \frac{1}{L}\sum_{ij} e^{-\frac{r_i^2+r_j'^2}{4\sigma^2}} I_1\left(\frac{r_i r'_j}{2\sigma^2}\right)\mathbf{R}_{ij}, \quad (24)$$

where $I_1(\cdot)$ is a modified Bessel function of the first kind. The kernel in (24) is rotation-covariant by construction as can be seen immediately by comparison with Eq. (7).

By exploiting the internal structure of the orthogonal group discussed above, it is straightforward to show that the rotoreflection covariant kernel is given by

$$\mathbf{K}^c_{O(2)}(\rho,\rho') = \tfrac{1}{2}\left[\mathbf{K}^c_{SO(2)}(\rho,\rho') + \mathbf{K}^c_{SO(2)}(\rho,\mathcal{Q}\rho')\mathbf{Q}\right], \quad (25)$$

which is the two-dimensional analog of Eq. (21). Interestingly, the resulting kernel can be also cast in the more intuitive form

$$\mathbf{K}^c_{O(2)}(\rho,\rho') = \frac{1}{L}\sum_{ij} e^{-\frac{r_i^2+r_j'^2}{4\sigma^2}} I_1\left(\frac{r_i r'_j}{2\sigma^2}\right)\hat{\mathbf{r}}_i\hat{\mathbf{r}}'^{\mathrm{T}}_j, \quad (26)$$

where the hat denotes a normalized vector. Equation (26) implies that the predicted force on an atom at the center of a configuration ρ will be a sum of pairwise forces oriented along the directions $\hat{\mathbf{r}}_i$ connecting the central atom with each of its neighbors (while each neighbor will experience a corresponding reaction force). The modulus of these forces will be a function of the interatomic distance completely determined by the training database, whose integral can be thought of as a pairwise energy potential. Clearly then, the resulting force field will be conservative: for any fixed database, the forces predicted by GP inference using this kernel will do zero work if integrated along any closed trajectory loop in configuration space.

To test the relative performance of the learning models discussed above, we constructed training and test databases for a two-dimensional triangular lattice, sampled from a constant temperature molecular dynamics simulation of a 48-particle system interacting via standard Lennard-Jones forces, once more using periodic boundary conditions and a Langevin thermostat. As the chosen lattice has threefold and sixfold symmetry, we can also examine the performance of covariant kernels that obey the two properties described above restricted to appropriate finite groups; these kernels are constructed as in Eq. (17). In this way we can monitor how imposing a progressively higher degree of symmetry on the kernel changes the rate at which forces in this system can be learned.

Our results are reported in Fig. 3. As anticipated, we find that the discrete covariant summation over the elements of a group G is approximately equivalent to a $|G|$-fold increase in the number of data points. This can be seen, e.g., from the results for the C_3 kernel (threefold rotations) and the C_6 kernel (sixfold rotations), by comparing the error incurred in the two cases using 20 and 10 data points, respectively. More generally, we observe that the larger the group, the faster the learning. Note, however, that for the covariant summation (17) to extract content from the database that is actually useful for predicting forces in the test configurations at hand, the group used must describe a true underlying point symmetry of the system. Hence, for instance, the C_4 kernel gives rise to much slower learning than the C_3 kernel for the 2D triangular lattice examined. Consistently, for this lattice the full point group D_6 performs almost as well as the continuous symmetry kernels, suggesting that not much more is to be gained once the full (finite-group) symmetry of a system has been captured. This finding enables accurate force prediction in a crystalline system when base kernels are used for which the covariant

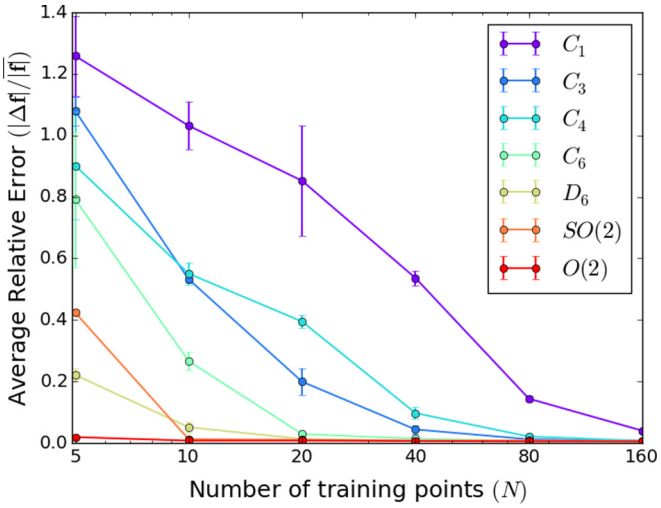

FIG. 3. Learning curves for 2D triangular grid of LJ atoms. The larger the symmetry group used to construct the kernel, the faster the learning, provided that the lattice symmetry is captured.

integration cannot be performed analytically, because the summation over a discrete symmetry group is available as a viable alternative.

C. 3D systems

We next benchmark the accuracy of our kernels in predicting DFT forces in three-dimensional bulk metal systems. As in the 2D case, starting from the linear base kernel we proceed to carry out the covariant integration analytically. After expressing the integration as a sum of pairwise integrals, the position vectors $\mathbf{r}_i$ and $\mathbf{r}'_j$ of two atoms in each pair are aligned onto each other. A convenient way to achieve this is by making both vectors parallel to the z axis with appropriate rotations $\mathbf{R}^z_i$ and $\mathbf{R}^z_j$. As before, the covariant integration will yield a matrix whose elements are scalar functions of the radii r_i and r'_j only. The integration can be carried out analytically over the standard three Euler angle variables (cf. Supplemental Material [42] for further details). Due to the z-axis orientation, the kernel matrix elements turn out to be all zero except for the zz one. The result reads

$$\begin{aligned}
\mathbf{K}^c_{SO(3)}(\rho,\rho') &= \frac{1}{L}\sum_{ij}\mathbf{R}^{z\mathrm{T}}_i\begin{pmatrix}0&0&0\\0&0&0\\0&0&\phi(r_i,r'_j)\end{pmatrix}\mathbf{R}^z_j,\\
\phi(r_i,r_j) &= \frac{e^{-\alpha_{ij}}}{\gamma^2_{ij}}(\gamma_{ij}\cosh\gamma_{ij}-\sinh\gamma_{ij}),\\
\alpha_{ij} &= \frac{r_i^2+r_j'^2}{4\sigma^2},\\
\gamma_{ij} &= \frac{r_i r'_j}{2\sigma^2}.
\end{aligned} \tag{27}$$

As in the 2D case, this covariant kernel matrix can be rewritten in terms of the unit vectors $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{r}}'_j$ associated with the atoms of the configurations ρ,ρ' as

$$\mathbf{K}^c_{SO(3)}(\rho,\rho') = \frac{1}{L}\sum_{ij}\phi(r_i,r'_j)\hat{\mathbf{r}}_i\hat{\mathbf{r}}'^{\mathrm{T}}_j, \tag{28}$$

making it apparent that the kernel models a pairwise conservative force field. However, while in 2D we needed to impose the full rotoreflection symmetry in order to obtain Eq. (26), rotations alone are sufficient to arrive at the fully covariant kernel in (28). This is a consequence of the fact that, in three dimensions, the covariant integral over rotations already imposes that the predicted force any atom will exert on any other is aligned along the vector connecting the pair: by symmetry there can be no preferred direction for an orthogonal force component after integrating over all rotations around the connecting vector, so that $\mathbf{K}^c_{O(3)} = \mathbf{K}^c_{SO(3)}$. This is not the case in two dimensions where covariant integration is over rotations around the z axis orthogonal to all connecting vectors lying in the xy plane, so that nonaligned predicted force components associated with a nonzero torque are not forbidden by symmetry in $\mathbf{K}^c_{SO(2)}$, and only the fully symmetrized kernel (25) will reduce to the pairwise form (26). More generally we may conjecture that the rotationally covariant kernel $\mathbf{K}^c_{O(d)}$ derived from a linear base kernel predicts pairwise central forces, and hence is conservative, in any dimension d.

We note that energy conserving kernels have previously been obtained as double derivatives (Hessian matrices) of scalar energy kernels (as originally described in Refs. [43,44] and used for atomistic systems in Ref. [38] to learn energies and more recently in Ref. [45] to learn forces). However, no closed-form expressions exist for the energy kernels that would yield our $O(d)$ energy conserving kernels through this route, since the required double integration of the kernels (21), (26), (28) cannot be carried out analytically.

To test our models, we performed DFT-accurate dynamical simulation with exchange and correlation energy modeled via the PBE/GGA approximation [46]. The systems considered were $4\times4\times4$ supercells of fcc nickel and bcc iron in periodic boundary conditions. A weakly coupled Langevin thermostat was used to control the temperature. We first examine bulk nickel at the target temperatures of 500 K and 1700 K, i.e., for an intermediate temperature where anharmonic behavior is already significant, and at a temperature close to the melting point where the strong thermal fluctuations make the system explore a more complex target configuration space. Figure 4 illustrates the performance of the kernel in Eq. (27) on this system.

The effect of adding symmetry information on the learning curve is very significant for both temperatures. In particular, the $SO(3)$ covariant kernel achieves a force error average lower than the 0.1 eV/Å threshold using remarkably few training points: 10 and 80 for the lower and higher temperatures in this test, respectively. The errors of the most accurate models (achieved with a $N = 320$ database) are particularly low: $0.0435(\pm0.0006)$ eV/Å and $0.095(\pm0.003)$ eV/Å respectively. Moreover, we note that the error on each force component (often reported in the literature, and different from the error on the full force vector used here) will be lower by a factor $\sqrt{3}$. This yields errors of 0.025 eV/Å and 0.052 eV/Å in the two cases, the former comparing well with the 0.09 eV/Å value obtained by using a state-of-the-art embedded atom model (EAM) interatomic potential for nickel [47,48].

Figure 5 allows one to assess the accuracy of the GP predictions in a complementary way: here we plot the probability distribution of the atomic forces as a function

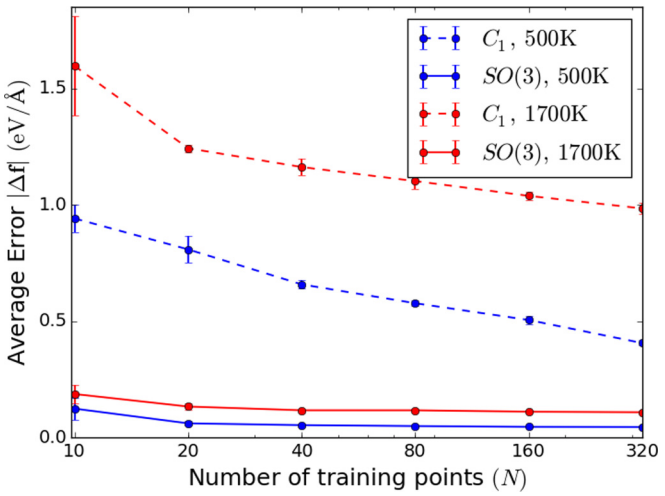

FIG. 4. Learning curves for crystalline nickel at two target temperatures. The $SO(3)$ covariant kernel (full lines) outperforms the base one (dashed lines).

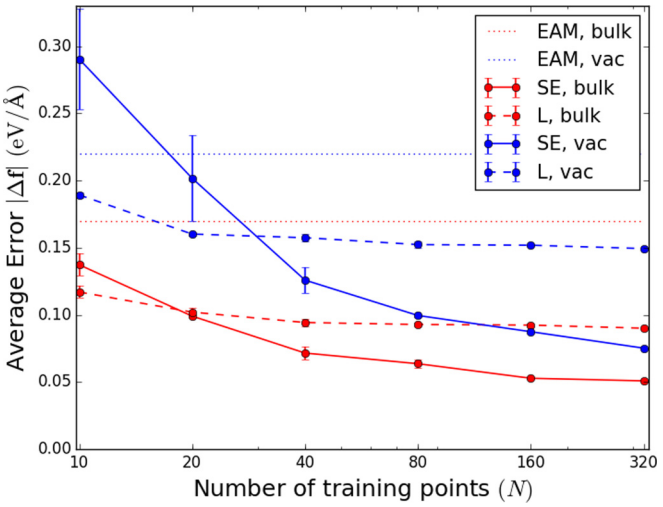

FIG. 6. Learning curves associated with force prediction by the linear (L, dashed lines) and squared exponential (SE, solid lines) covariant kernels in bulk iron systems. Red and blue colors indicate undefected systems and model systems containing a vacancy, respectively.

of the force modulus (blue histogram) and the associated relative error density (gray histogram). We define the latter as $RED(f) = \frac{|\Delta \mathbf{f}|}{f} p(f)$, which is normalized to 0.055, reflecting the 5.5% average relative error incurred by force prediction. The fact that $RED(f)$ is everywhere a small fraction of $p(f)$ demonstrates that a reasonable accuracy is achieved for the whole range of forces predicted.

The results presented so far indicate that fully exploiting symmetry significantly improves the accuracy of force prediction. Covariance is thus always used in the following analysis, where we compare the performance of different symmetry-aware kernels. We start by choosing iron systems for these tests as many properties of iron-based systems remain out of modeling reach. This is mostly due to technical limitations. On the one hand, full DFT calculations on large systems are too computationally expensive and even hybrid quantum-classical (QM/MM) simulations of iron systems are typically overwhelmingly costly, as they require large QM-zone buffered clusters to fully converge the forces [49]. On the other hand, in many situations even the best available state-of-the-art classical force fields may not guarantee accurate force prediction, as they may incur systematic errors [48,49], or may be hard to extend to complex chemical compositions [50], so that a technique that can indefinitely reuse all computed QM forces via GP inference and produce results that are traceably aligned with DFT-accurate forces could be very useful [26,51].

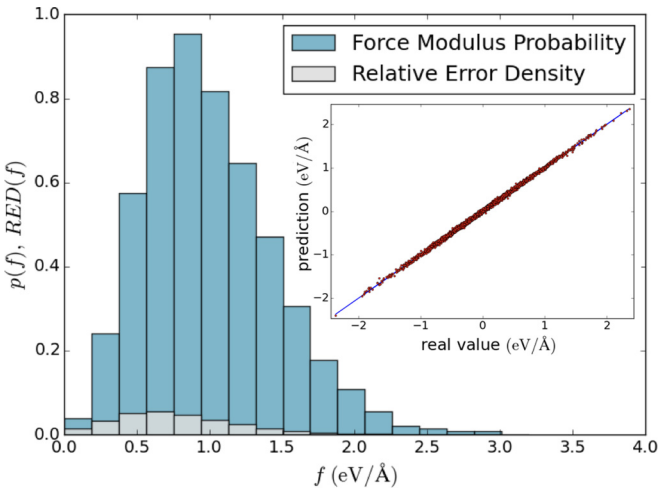

FIG. 5. Density of relative error made by the GP algorithm ($N = 320$) for bulk nickel at 500 K. The inset shows the scatter plot of real vs. predicted cartesian components for the same data.

We carried out constant temperature (500 K) molecular dynamics simulations of two bcc iron systems: a 64-atom crystalline system and a 63-atom system derived from this and containing a single vacancy. In the latter, only the atoms within the first two neighbor shells of the vacancy were used to test the algorithm, to better resolve the performance of our kernels in a defective system. Figure 6 shows the learning curves for the two symmetrized kernels: the linear kernel covariant over $O(3)$ and the squared exponential kernel (20) covariant over the full cubic point group of the crystal. The figure also reports the performance of a high-quality EAM potential [52]. Both kernels perform better than the EAM potentials in this test. However, the error rate of the linear kernel (dashed lines) levels off to some constant nonzero value that might or might not be satisfactory (depending on the application), and will generally depend on the system being examined. In bulk iron the error floor value is about 0.09 eV/Å while in the vicinity of a vacancy it is considerably higher (0.15 eV/Å), suggesting that in spite of its many attractive properties (e.g., fast evaluation, fast convergence, energy conservation), the linear class of kernels of the form (28) is by no means complete, that is, it sometimes cannot capture and reproduce the entirety of the reference QM physical interaction. In many situations, kernels capable of reproducing higher-order interactions could be needed to reach the target accuracy. This is exemplified by the much better performance of the squared exponential kernel (full lines in the figure), which yields higher

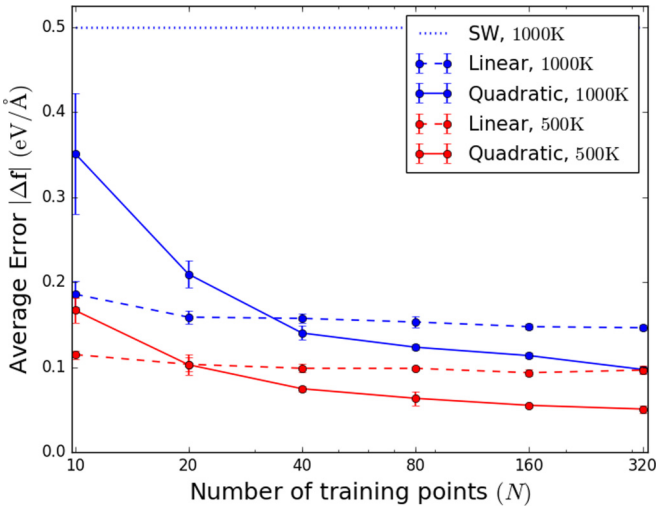

FIG. 7. Learning curves obtained for crystalline silicon using the linear kernel (dashed lines) or the quadratic kernel (solid lines). Different colors indicate different temperatures.

accuracy, particularly for the more complex vacancy system (about 0.05 eV/Å and 0.075 eV/Å for atoms in the bulk and near the vacancy respectively). It is worth noting here that, in general, conserving energy exactly by construction provides no guarantee of higher force accuracy. For instance, in the case above, the squared exponential kernel delivers much more precise forces even though it conserves energy only approximately. As the approximation will in any case improve with the accuracy of the predicted forces, while no $SO(3)$-invariant energy conserving equivalent of this kernel has been proposed or appears viable, whether it is preferable to use this kernel or a less accurate but energy conserving alternative one, will generally depend on both the target system and the application at hand.

For target systems with no clear point symmetry, a full covariant integration would always be desirable. This cannot be carried out analytically for the squared exponential kernel, where symmetrizing by a discrete summation is the only option. However, interactions beyond pairwise can be still captured by the quadratic kernel obtained by taking the square of the linear kernel (19). In contrast to the squared exponential kernel, this is analytically tractable (for instance, an $SO(3)$-invariant scalar quadratic kernel was obtained in Ref. [37]), and our analysis reveals that a matrix-valued quadratic kernel covariant over $O(3)$ can be derived analytically (details of the calculation are a subject for future work [53]). The resulting model generates a rotoreflection symmetric three-body force field that can be expected to properly describe non-close-packed bonding, such as found in covalent systems, for example.

Figure 7 illustrates the errors incurred by the linear and the quadratic kernel while attempting to reproduce the forces obtained during Langevin dynamics of a 64-atom crystalline silicon system using density functional tight binding (DFTB) [54]. Both linear and quadratic kernels are significantly more accurate than a classical Stillinger-Weber (SW) potential [55] fitted to reproduce the DFTB lattice parameter and bulk modulus [26]. Due to its more restricted associated function space, the linear kernel is the one that learns faster, and would be the more accurate if only very restricted databases had to be used. However, the quadratic kernel eventually performs much better than the (effective two-body) linear one for both of the temperatures, 500 K and 1000 K, that we investigated in this covalent system. We obtain errors of 0.05 eV/Å and 0.1 eV/Å in the two cases, corresponding, respectively, to approximately 4% and 6% of the mean force. These are very close to the minimum baseline locality error [56] associated with the finite cutoff radius used for the Gaussian expansion in (18).

V. CONCLUSION

In this work we presented a new method to learn quantum forces on local configurations. This method is based on a vectorial Gaussian process that encodes prior knowledge in a matrix-valued kernel function. We showed how to include rotation and reflection symmetry of the force in the GP process via the notion and use of covariant kernels. A general recipe was provided to impose this property on otherwise nonsymmetric kernels. The essence of this recipe lies in a special integration step, which we call covariant integration, over the full rotoreflection group associated with the relevant number of system dimensions. This calculation can be performed analytically starting from a linear base kernel, and the resulting $O(d)$ covariant kernels can be shown to generate conservative force fields.

We furthermore tested covariant kernels on standard physical systems in one, two, and three dimensions. The one- and two-dimensional scenarios served as playgrounds to better understand and illustrate the essential features of such learning. The 3D systems allowed some practical benchmarking of the methodology in real systems. In agreement with what physical intuition would suggest, we consistently found that incorporating symmetry gives rise to more efficient learning. In particular, if both database and target configurations belong to a system with a definite underlying symmetry, restricting kernel covariance to the corresponding finite symmetry group will deliver the full speed-up of error convergence with respect to database size. At the same time this approach lifts the requirement of analytical integrability over the full $SO(d)$ manifold, as the restricted integration becomes a simple discrete sum over the relevant finite set of group elements. Testing on nickel, silicon, and iron (the latter both pure and defective) reveals that the present recipes can improve significantly on available classical potentials. In general, nonlinear kernels may be needed for accurate force predictions in the presence of complicated interactions, e.g., in the study of plasticity or embrittlement/fracture behavior of covalent or metallic systems. In particular, a quadratic base kernel yields a fully $O(3)$ covariant effective three-body force field, and our tests suggest that this can be used successfully to improve the accuracy of force prediction in covalent materials. Current work is focusing on amorphous Si systems, where the lack of a clear point symmetry makes the full $O(3)$ covariance strictly necessary.

Our results reveal that force covariance is achievable without imposing energy conservation to the kernel form. While both are desirable properties, we find that lifting the exact energy conservation constraint can sometimes yield

higher force accuracy. For instance, no invariant local energy based kernel has been proposed for the squared exponential (universal approximator) kernel, since the analytic integration over $SO(3)$ is not viable. However, we find that covariance limited to the O_{48} point group is very effective for force predictions in crystalline Fe systems using this kernel (see Fig. 6).

In general, while predicting forces with high accuracy is the main motivation for machine learning-based work in this field, the best compromise between accuracy, energy conservation and covariance will depend on the specific target application. For instance, kernels built from a covariant integration (or summation) that do not conserve energy exactly should not be used as substitutes for conventional interatomic potentials to perform long NVE simulations, since they might in principle lead to non-negligible spurious energy drifts. This is not a problem in NVT simulations, where a thermostat exchanges energy with the system to achieve and conserve the target temperature, which will be able to compensate for any such drift if appropriately chosen [57]. Furthermore, the same kernels will be particularly suited for schemes that are in all cases incompatible with strict energy conservation. These include the LOTF approach and any online learning scheme similarly involving a dynamically updated force model. They also include any highly accurate and transferable scheme based on a fixed, very large database where, to maximize efficiency, each force prediction only uses its corresponding most relevant database subset.

On the other hand, any usage style is possible for covariant kernels conserving energy exactly, such as the covariant linear kernels of Eqs. (21), (26), and (28). In fact, the conservative pairwise interaction forces generated by these covariant linear kernels can be easily integrated to provide effective optimal standard pairwise potentials for any application needing a total energy expression. We also note that while the pair interaction form would still ensure very fast evaluation of the predicted forces, its accuracy for complex systems could be improved by dropping the transferability requirement of a single pairwise function. In such a scheme, different system regions could conceivably be modeled by locally optimized forces/potentials, where the local tuning could be simply achieved by restricting the inference process to subsets of the database pertinent to each target region.

ACKNOWLEDGMENTS

The authors acknowledge funding by the Engineering and Physical Sciences Research Council (EPSRC) through the Centre for Doctoral Training "Cross Disciplinary Approaches to Non-Equilibrium Systems" (CANES, Grant No. EP/L015854/1), by the Office of Naval Research Global (ONRG Award No. N62909-15-1-N079). A.D.V. acknowledges further support by the EPSRC HEmS Grant No. EP/L014742/1 and by the European Union's Horizon 2020 research and innovation program (Grant No. 676580, The NOMAD Laboratory, a European Centre of Excellence). The research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-06CH11357. A.G. would like to thank F. Bianchini for his help in data collection and R. G. Margiotta, K. Rossi, F. Bianchini, and C. Zeni for useful discussions.

[1] G. E. Moore, Cramming more components onto integrated circuits, Proc. IEEE **86**, 82 (1998).
[2] C. Walter, Kryder's Law, Sci. Am. **293**, 32 (2005).
[3] E. Grochowski and R. D. Halem, Technological impact of magnetic hard disk drives on storage systems, IBM Syst. J. **42**, 338 (2003).
[4] D. R. Hartree, The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods, Math. Proc. Cambridge **24**, 89 (1928).
[5] V. Fock, Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, Zeitschrift für Physik **61**, 126 (1930).
[6] J. C. Slater, A simplification of the Hartree-Fock method, Phys. Rev. **81**, 385 (1951).
[7] Nomad Repository, https://repository.nomad-coe.eu/
[8] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, Big Data of Materials Science: Critical Role of the Descriptor, Phys. Rev. Lett. **114**, 105503 (2015).
[9] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, Accelerating materials property predictions using machine learning, Sci. Rep. **3**, 1 (2013).
[10] C. Kim, G. Pilania, and R. Ramprasad, From organized high-throughput data to phenomenological theory using machine learning: The example of dielectric breakdown, Chem. Mater. **28**, 1304 (2016).
[11] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. von Lilienfeld, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett. **108**, 058301 (2012).
[12] J. C. Snyder, M. Rupp, K. Hansen, K. R. Müller, and K. Burke, Finding Density Functionals with Machine Learning, Phys. Rev. Lett. **108**, 253002 (2012).
[13] L. F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, Machine learning for many-body physics: The case of the Anderson impurity model, Phys. Rev. B **90**, 155136 (2014).
[14] A. Lopez-Bezanilla and O. A. von Lilienfeld, Modeling electronic quantum transport with machine learning, Phys. Rev. B **89**, 235411 (2014).
[15] 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**, 136403 (2010).
[16] J. Behler and M. Parrinello, Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, Phys. Rev. Lett. **98**, 146401 (2007).
[17] A. V. Shapeev Moment tensor potentials: A class of systematically improvable interatomic potentials, Multiscale Model. Simul. **14**, 1153 (2016).
[18] T. Stecher, N. Bernstein, and G. Csányi, Free energy surface reconstruction from umbrella samples using gaussian process regression, J. Chem. Theor. Comput. **10**, 4079 (2014).

[19] A. De Vita and R. Car, A novel scheme for accurate Md simulations of large systems, MRS Proc. **491**, 473 (1997).

[20] G. Csányi, T. Albaret, M. C. Payne, and A. De Vita, "Learn on the Fly": A Hybrid Classical and Quantum-Mechanical Molecular Dynamics Simulation, Phys. Rev. Lett. **93**, 175503 (2004).

[21] E. V. Podryabinkin and A. V. Shapeev, arXiv:1611.09346.

[22] B. Haasdonk and H. Burkhardt, Invariant kernel functions for pattern analysis and machine learning, Mach. Learn. **68**, 35 (2007).

[23] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, 2006).

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics Series (Springer, New York, 2006).

[25] M. Krejnik and A. Tyutin, Reproducing kernel hilbert spaces with odd kernels in price prediction, IEEE T. Neur. Net. Lear. **23**, 1564 (2012).

[26] 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**, 096405 (2015).

[27] M. Caccin, Z. Li, J. R. Kermode, and A. De Vita, A framework for machine-learning-augmented multiscale atomistic simulations on parallel supercomputers, Int. J. Quant. Chem. **115**, 1129 (2015).

[28] V. Botu and R. Ramprasad, Adaptive machine learning framework to accelerate *ab initio* molecular dynamics, Int. J. Quant. Chem. **115**, 1074 (2014).

[29] C. A. Micchelli and M. Pontil, Kernels for multi-task learning, in *Advances in Neural Information Processing Systems* (University at Albany Press, Albany, 2005).

[30] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, Kernels for vector-valued functions: A review, Foundations and Trends® in Machine Learning (now publishers inc., Boston, 2012), Vol. 4, pp. 195–266.

[31] The DFT datasets used in this work are openly available from the research data management system of King's College London at http://doi.org/10.18742/RDM01-92.

[32] W. Kohn, Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms, Phys. Rev. Lett. **76**, 3168 (1996).

[33] E. Prodan and W. Kohn, Nearsightedness of electronic matter, Proc. Natl. Acad. Sci. USA **102**, 11635 (2005).

[34] T. Bayes and M. Price, An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, frs communicated by Mr. Price, in a letter to John Canton, amfrs, Philos. Trans. (1683–1775), 1763.

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

[36] V. Botu and R. Ramprasad, Learning scheme to predict atomic forces and accelerate materials simulations, Phys. Rev. B **92**, 094306 (2015).

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

[38] A. P. Bartók and G. Csányi, Gaussian approximation potentials: A brief tutorial introduction, Int. J. Quantum Chem. **115**, 1051 (2015).

[39] M. L. Mehta, *Random Matrices*, 3rd ed., Pure and applied mathematics series (Elsevier, San Diego, 2004).

[40] S. Aubert and C. S. Lam, Invariant integration over the unitary group, J. Math. Phys. **44**, 6112 (2003).

[41] J. Mercer, Functions of positive and negative type, and their connection with the theory of integral equations, Proc. Roy. Soc. Lond. A-Conta. **83**, 69 (1909).

[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.95.214302 for further details on the covariant integration of the linear kernels.

[43] E. J. Fuselier Jr., Ph.D. thesis, Texas A and M University, 2006.

[44] I. Macêdo and R. Castro, *Learning Divergence-Free and Curl-Free Vector Fields with Matrix-Valued Kernels* (Technical Report, Instituto Nacional de Matematica Pura e Aplicada, Rio de Janeiro, 2008).

[45] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller Machine learning of accurate energy-conserving molecular force fields, Sci. Adv. **3**, e1603015 (2017).

[46] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. **77**, 3865 (1996).

[47] Y. Mishin, Atomistic modeling of the γ and γ'-phases of the Ni–Al system, Acta Mater. **52**, 1451 (2004).

[48] F. Bianchini, J. R. Kermode, and A. De Vita, Modelling defects in Ni–Al with EAM and DFT calculations, Model. Simul. Mater. Sci. **24**, 045012 (2016).

[49] F. Bianchini, A. Glielmo, J. R. Kermode, and A. De Vita (unpublished).

[50] J. von Pezold, L. Lymperakis, and J. Neugebeauer, Hydrogen-enhanced local plasticity at dilute bulk H concentrations: The role of H-H interactions and the formation of local hydrides, Acta Mater. **59**, 2969 (2011).

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

[52] M. I. Mendelev, S. Han, D. J. Srolovitz, G. J. Ackland, D. Y. Sun, and M. Asta, Development of new interatomic potentials appropriate for crystalline and liquid iron, Philos. Mag. **83**, 3977 (2003).

[53] A. Glielmo, P. Sollich, and A. De Vita (unpublished).

[54] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, Phys. Rev. B **58**, 7260 (1998).

[55] F. H. Stillinger and T. A. Weber, Computer simulation of local order in condensed phases of silicon, Phys. Rev. B **31**, 5262 (1985).

[56] V. L. Deringer and G. Csányi, Machine learning based interatomic potential for amorphous carbon, Phys. Rev. B **95**, 094203 (2017).

[57] A. Jones and B. Leimkuhler, Adaptive stochastic methods for sampling driven molecular systems, J. Chem. Phys. **135**, 084125 (2011).