Accurate representation of formation energies of crystalline alloys with many components

A. Shapeev

Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Nobel Str. 3, Moscow 143026, Russia

ABSTRACT

In this paper I propose a new model for representing the formation energies of multicomponent crystalline alloys as a function of atom types. In the cases when displacements of atoms from their equilibrium positions are not large, the proposed method has a similar accuracy as the state-of-the-art cluster expansion method, and a better accuracy when the fitting dataset size is small. The proposed model has only two tunable parameters—one for the interaction range and one for the interaction complexity.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Accurate computational prediction of properties of alloys from their composition is one of the outstanding problems of rational materials design. Emerging applications, such as the high-entropy alloys (HEAs) [1], pose new challenges to computational materials methods. HEAs are defined as alloys with five or more constituent elements in equal or close to equal proportions [1]. A large number of elements leads to high configuration entropy of the solid solution phase and hence stabilizes it. Owing to this, HEAs possess many unique mechanical properties [2–4]. Accurate computational prediction of the mixing enthalpy and configuration entropy would be very instrumental in studying HEAs, as it is hard to experimentally explore different compositions of five or more elements due to combinatorial complexity. The state-of-the-art methodology of computationally assessing the stability of multicomponent crystalline alloys is based on cluster expansion [5–8], allowing to fit formation energies of binary systems over the entire range of compositions, ternary and quaternary systems [9–12] over, typically, some subrange of the composition range, and quinary systems at specific points of the composition range [13].

Cluster expansion belongs to a class of data-driven interatomic interaction models, along with machine learning interatomic potentials [14–17]. Data-driven models assume a flexible functional form of the interaction energy with many (hundred or more) free parameters that are fitted from the quantum-mechanical data.

The major drawback of cluster expansion is the "relaxation error" [8] which refers to low ability of cluster expansion to account for changes in mixing enthalpy related to the displacements of atoms to their equilibrium positions due to, for instance, the mismatch in their atomic radii. This is probably the reason why cluster expansion is not very successful in handling systems with a large number of components.

In this paper I propose a new approach to accurately representing and fitting the formation energies of alloys with a large number of elements. The new approach does not directly resolve the "relaxation error" issue, but is shown to have a comparable or better performance than cluster expansion in cases when the "relaxation error" is low. The approach is based on partitioning the energy into contributions of the atomic environments and representing these contributions with low-rank multidimensional tensors [18]. The proposed model has only two adjustable parameters, the range of the interatomic interaction and the upper bound on the tensor rank, the latter controls the number of free parameters in the model. The idea of a low-rank representation of functions of the atomic environments was also pursued in [19].

2. Interatomic interaction

I consider the following model of a crystal. Let the undeformed positions of atoms be $A\mathbb{Z}^3$, where $\mathbb{Z}^3$ is the lattice of all points with integer coordinates and the matrix A sets the actual crystal structure and dimensions. For example, a face-centered cubic (f.c.c.)

E-mail addresses: alexander@shapeev.com, a.shapeev@skoltech.ru

lattice with constant a is defined by $A = \begin{pmatrix} 0 & a/2 & a/2 \\ a/2 & 0 & a/2 \\ a/2 & a/2 & 0 \end{pmatrix}$.

Each atom $\xi \in A\mathbb{Z}^3$ is displaced by $x(\xi) \in \mathbb{R}^3$ from its reference position ξ and is of the type $\sigma(\xi) \in \{1, \ldots, m\}$.[1] Let $\Omega \subset A\mathbb{Z}^3$ be a computational domain (supercell) repeated periodically in the entire space. Then the degrees of freedom of the atomistic system are $x = x(\xi)$ and $\sigma = \sigma(\xi)$ for each $\xi \in \Omega$ and the interaction energy is $\widetilde{E} = \widetilde{E}(x, \sigma)$. I will refer to this as an *on-lattice* model of a crystal, since the atoms are indexed by their undeformed position in an ideal lattice. This is in contrast with, for example, interatomic potentials, such as the embedded atom method (EAM) or the machine learning based ones [14–17], that do not relate atoms to their undeformed positions.

Unless we want to model defects, the displacements x can be eliminated from the model. Often, it is assumed that the atoms are at (or near) their relaxed positions and hence

$$E(\sigma) := \min_x \widetilde{E}(x, \sigma). \tag{1}$$

This model is good for sampling different distributions of atoms to the lattice sites and hence, e.g., estimating the configuration entropy of an alloy with a known lattice (for which the atomic displacements are not dramatic).

The atoms are assumed to interact only with their closest environment characterized by the cut-off distance R_{cut} and the interaction neighborhood comprised of all vectors $r_1, \ldots, r_n \in A\mathbb{Z}^3$ whose length is less than R_{cut}. Hence, the energy of interaction of these atoms is postulated to be

$$E(\sigma) = \sum_{\xi \in \Omega} V(\sigma(\xi + r_1), \ldots, \sigma(\xi + r_n)), \tag{2}$$

where V is called, by analogy with the off-lattice case, the "interatomic potential". Essentially, V is an $m \times m \times \ldots \times m$, n-dimensional tensor that defines the interaction model (2). The goal is to fit this model to the "true" quantum-mechanical model given by $E^{\text{qm}}(\sigma)$. Without loss of generality it can be assumed that $E^{\text{qm}}(\sigma) = 0$ whenever all $\sigma(\xi)$ are equal—in this case $E^{\text{qm}}(\sigma)$ is called the formation energy of σ.

The fitting is done on a set of K atomistic configurations $\sigma^{(k)}, k = 1, \ldots, K$, given together with their true energies $E^{\text{qm}}(\sigma^{(k)})$. Then, the sought V is obtained by minimizing the mean-square functional

$$\frac{1}{K} \sum_{k=1}^{K} \left| E(\sigma^{(k)}) - E^{qm}(\sigma^{(k)}) \right|^2. \tag{3}$$

This is a linear regression problem on the multidimensional tensor V.

The typical datasets quoted in the literature have a few hundreds of configurations, whereas, for example, for an f.c.c. crystal even for two species ($m = 2$) and 12 nearest neighbors ($n = 13$) the problem (3) has $m^n = 8192$ unknown parameters fitting which already seems intractable. Accounting, however, for physical symmetries (i.e., enforcing V to be invariant with respect to the f.c.c. crystal symmetry space group) reduces the number of unknowns to 288 and the problem becomes tractable. However, if the number of species increases to $m \geqslant 3$ then the number of unknowns becomes of the order of 10^5 or more, which necessitates further reduction in the number of unknowns. Below I show that the so-called low-rank tensor representation for V successfully reduces the number of unknowns in the regression problem and yields an efficient way of accurately fitting the formation energies.

---

[1] In the cluster expansion literature the degrees of freedom are typically denoted by $\sigma(\xi)$, following the notation of the electron spin.

3. Formation energy representation

3.1. Low-rank Tensors

I first review the concept of low-rank matrices and tensors. Consider an $m \times m$ matrix $M = M(i, j)$, where $i, j \in \{1, \ldots, m\}$. The matrix has rank r or less if it can be represented as

$$M(i, j) = \sum_{\ell=1}^{r} u_\ell(i)\, v_\ell(j),$$

where u and v are the scaled singular vectors of M. As a natural generalization, we can say that a tensor V has rank r if

$$V(\sigma_1, \ldots, \sigma_n) = \sum_{\ell=1}^{r} u_\ell^{(1)}(\sigma_1), \ldots, u_\ell^{(n)}(\sigma_n) \tag{4}$$

for some $u_\ell^{(1)}, \ldots, u_\ell^{(n)}$. It is known that the set of tensors of the form (4) is not a closed set, i.e., a limit of a sequence of low-rank tensors may be a high-rank tensor. This can be illustrated, for example, by taking a simple one-body energy $V(\sigma_1, \ldots, \sigma_n) = \varphi(\sigma_1) + \cdots + \varphi(\sigma_n)$ which is a rank-n tensor, but it can be approximated with an arbitrary accuracy by the following rank-two tensor:

$$\frac{(1 + \epsilon\varphi(\sigma_1)), \ldots, (1 + \epsilon\varphi(\sigma_n)) - 1}{\epsilon} \approx (\varphi(\sigma_1) + \ldots + \varphi(\sigma_n)).$$

Due to this, the low-rank tensors might not behave well when iteratively solving an optimization problem.

I will therefore use an alternative version of low-rank tensors that is free from this problem. Namely, I will use the so-called tensor train (TT) representation [18] defined by

$$V(\sigma_1, \ldots, \sigma_n) = A^{(1)}(\sigma_1), \ldots, A^{(n)}(\sigma_n), \tag{5}$$

where each $A^{(i)} = A^{(i)}(\sigma_i)$ is an σ_i-dependent matrix of size $r_{i-1} \times r_i$, and $r_0 = r_n = 1$. The rank of this representation is defined as $\bar{r} := \max_i r_i$. I will call the model (5) the low-rank potential (or LRP). The LRP has about $nm\bar{r}^2$ free parameters, which is much less than m^n as was before using the low-rank assumption.

3.2. Fitting the interatomic potential

I thus assume that the interatomic potential V has the form (5), where, after R_{cut} is fixed, $\bar{r}$ is the only parameter controlling the accuracy of the model. The σ-dependent matrices $A^{(i)}$ are the unknown parameters of the model that are found by fitting to the ab initio data. Totally, there is $O(nm\bar{r}^2)$ parameters. It should be noted that (5) is not invariant with respect to the space group of the cubic lattice, G, consisting of 48 linear space transformations $g \in \mathbb{R}^{3 \times 3}$ that map $A\mathbb{Z}^3$ into itself. The mean-square functional (3) is therefore adjusted to

$$J := \frac{1}{48K} \sum_{g \in G} \sum_{k=1}^{K} \left| E(g(\sigma^{(k)})) - E^{qm}(\sigma^{(k)}) \right|^2. \tag{6}$$

The training (fitting) problem is hence:

find $A^{(1)}, \ldots, A^{(n)}$ minimizing J subject to (2) and (5).

To solve this optimization problem, it should be noted that the functional J is quadratic in E, the latter is linear in V, and V is linear in each $A^{(i)}$. I hence use the alternating least squares (ALS) algorithm (see, e.g., [20]), consisting of taking an initial guess for all $A^{(i)}$ and then updating $A^{(i)}$ in an iterative manner until convergence. Each iteration consists of n subiterations, where in the i-th subiteration J is minimized with respect to $A^{(i)}$ while freezing $A^{(j)}, j \neq i$. The latter is a standard quadratic optimization problem which is