

UC Berkeley

UC Berkeley Previously Published Works

Title

An L_0L_1 -norm compressive sensing paradigm for the construction of sparse predictive lattice models using mixed integer quadratic programming

Permalink

<https://escholarship.org/uc/item/56v9s5t4>

Authors

Huang, Wenxuan
Urban, Alexander
Xiao, Penghao
et al.

Publication Date

2018-07-16

Peer reviewed

An L_0L_1 -norm compressive sensing paradigm for the construction of sparse predictive lattice models using mixed integer quadratic programming

Wenxuan Huang^{1,3}, Alexander Urban², Penghao Xiao³, Ziqin Rong^{1,3}, Hena Das², Tina Chen², Nongnuch Artrith², Alexandra Toumar^{1,3}, Gerbrand Ceder^{1,2,3}

1, Department of Material Science and Engineering, Massachusetts Institute of Technology, MA, USA

2, Department of Materials Science and Engineering, UC Berkeley, Berkeley, CA, USA

3, Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract

First-principles based lattice models allow the modeling of *ab initio* thermodynamics of crystalline mixtures for applications such as the construction of phase diagrams and the identification of ground state atomic orderings. The recent development of compressive sensing approaches for the construction of lattice models has further enabled the systematic construction of sparse physical models without the need for human intuition other than requiring the compactness of effective cluster interactions. However, conventional compressive sensing based on L_1 -norm regularization is strictly only applicable to certain classes of optimization problems and is otherwise not guaranteed to generate optimally sparse and transferable results, so that the method can only be applied to some materials science applications. In this paper, we illustrate a more robust L_0L_1 -norm compressive-sensing method that removes the limitations of conventional compressive sensing and generally results in sparser lattice models that are at least as predictive as those obtained from L_1 -norm compressive sensing. Apart from the

theory, a practical implementation based on state-of-the-art mixed-integer quadratic programming (MIQP) is proposed. The robustness of our methodology is illustrated for four different transition-metal oxides with relevance as battery cathode materials: $\text{Li}_{2x}\text{Ti}_{2(1-x)}\text{O}_2$, $\text{Li}_{2x}\text{Ni}_{2y}\text{O}_2$, $\text{Mg}_x\text{Cr}_2\text{O}_4$, and Na_xCrO_2 . This method provides a practical and robust approach for the construction of sparser and more predictive lattice models, improving on the compressive sensing paradigm and making it applicable to a much broader range of applications.

Introduction

First principles electronic density-functional theory (DFT) [1,2] has enabled predictive computational material science [3-9] and serves as an important tool for discovering novel materials [3,4,10]. However, the intrinsic computational scaling of DFT limits its applicability in practice to structures with a few hundred atoms and typically prevents the extensive configurational sampling that is required to examine temperature-dependent properties. To overcome this limitation, *lattice models* based on previously calculated DFT energies, such as *cluster expansion* (CE) or generalized Ising models [11-16], can be used to efficiently approximate first-principles energies of lattice configurations that are out of reach for DFT calculations.

As the name suggests, CE models approximate the energies of lattice configuration as a linear expansion in the occupancy (*effective cluster interactions*, ECIs) of site clusters, *i.e.*, site pairs, triplets, quadruplets, and so on. Typically, a small number of short-ranged clusters with few sites is sufficient to capture the characteristic physics of a material. However, it is not normally known *a priori* which clusters correspond to the most important interactions. Constructing a cluster expansion model, therefore, involves the solution of an underdetermined linear system in the cluster basis, with the objective of creating an optimal solution which is sparse, so

that most ECIs are zero and only the ECIs of the characteristic clusters contribute to the CE energy.

Traditionally, the characteristic clusters for a material were determined either based on human intuition [17] or using global optimization techniques, such as genetic algorithms [18]. Recently, important progress in CE model construction has been made [19-23] by making use of compressive sensing [24], a signal reconstruction technique that finds the sparsest solution for a certain type of underdetermined linear systems. The compressive-sensing CE approach allows the systematic and robust determination of sparse ECI parameters without the need for human intuition other than requiring the compactness of ECIs. However, the conventional L_1 -norm compressive sensing is only guaranteed to determine the optimally sparse solution of linear systems that fulfill a number of conditions, in particular the restricted isometry property (RIP) [25]. In practical material science and physics applications, these properties do not necessarily hold, and applying L_1 -norm compressive sensing does not automatically lead to the optimal model.

In this paper, we explore a generalization of the conventional L_1 -norm compressive sensing tailored for the domain of practical materials science CE problems that is not bound to the stringent requirements of the original method. We demonstrate that our L_0L_1 -norm compressive sensing method leads to significantly sparser and more predictive models.

The paper is structured as follows: First, we revisit the CE formalism and introduce the ECI fitting problem mathematically in the *methodology* section. Then, we introduce the L_0L_1 -norm compressive sensing paradigm and its mixed integer programming formulations, which enable efficient and near optimal solution of the relevant *nondeterministic polynomial time* (NP) hard problems with state of the art performance. Finally, in the *results* section we demonstrate that the L_0L_1 -norm formulation enables significantly sparser and even more predictive CE models compared to the conventional L_1 -norm formulation.

As a side note, ensuring the preservation of ground states is another important challenge in lattice model construction which has been addressed in reference [26]. The present paper focuses exclusively on constructing sparser lattice models but the formalism is compatible with that of reference [26] and should be understood as complementary

Methodology

Cluster expansion and compressive sensing

In this section, we briefly recap the concepts of cluster expansion. For more elaborate introductions, the readers are referred to the literature [11,14,19,27].

The general expression of a cluster expansion Hamiltonian is

$$\mathbf{E}_{\text{CE}}(\boldsymbol{\sigma}) = \left\langle J_0 + \sum_i^{\text{sites}} J_i \sigma_i + \sum_{i,j}^{\text{pairs}} J_{i,j} \sigma_i \sigma_j + \sum_{i,j,k}^{\text{triplets}} J_{i,j,k} \sigma_i \sigma_j \sigma_k + \dots \right\rangle = \sum_{c \in \mathbf{C}} J_c \langle \sigma_c \rangle = \boldsymbol{\Pi}(\boldsymbol{\sigma}) \mathbf{J}, \quad (1)$$

where $\boldsymbol{\sigma}$ is the lattice (*spin*) configuration with σ_i denoting the occupancy of site i . Conventionally, σ_i is +1 when the site is *spin up* in a magnetic system and -1 when the site is *spin down* though the variables can denote any binary occupancy. The polynomial $\sigma_i \sigma_j \dots$, denotes the *spin product* of a *cluster* of lattice sites, associated with the weight factor $J_{i,j,\dots}$, the *effective cluster interaction* (ECI). The CE energy is then the sum of the spin products multiplied with their corresponding ECIs and averaged over all sites within a lattice, as denoted by the angle brackets $\langle \rangle$. To simplify notation, with \mathbf{C} defined as the set of clusters, for any $c \in \mathbf{C}$, J_c and σ_c are shorthand notations of the previously defined ECIs $J_{i,j,\dots}$ and spin products $\sigma_i \sigma_j \dots$, respectively. The average of the spin products $\langle \sigma_c \rangle$ is the *cluster correlation* of cluster c . By concatenating the cluster correlations $\langle \sigma_c \rangle$ into a row vector $\boldsymbol{\Pi}(\boldsymbol{\sigma})$ and the ECIs J_c into a column vector \mathbf{J} , the CE energy of the lattice configuration $\boldsymbol{\sigma}$ can be simply written as the vector product $\boldsymbol{\Pi}(\boldsymbol{\sigma}) \mathbf{J}$.

Naïvely, the CE fitting problem can be formulated as follows: Given a set of input configurations S and their energies from DFT calculations $\mathbf{E}_{\text{DFT},S}$, minimize the root mean square error (RMSE) between the CE energy and DFT energy in the ECIs space, *i.e.*, $\min_{\mathbf{J}} \|\mathbf{E}_{\text{DFT},S} - \mathbf{E}_{\text{CE},S}\|_2$. By concatenating the correlation vectors of all configurations $\sigma \in S$ we obtain the correlation matrix $\mathbf{\Pi}_S$, which allows us to write more explicitly $\mathbf{E}_{\text{CE},S} = \mathbf{\Pi}_S \mathbf{J}$, so that the naïve statement of the CE fit problem becomes:

$$\min_{\mathbf{J}} \|\mathbf{E}_{\text{DFT},S} - \mathbf{\Pi}_S \mathbf{J}\|_2^2 . \quad (2)$$

For $p \geq 1$, the L_p norm of a vector \mathbf{u} is defined as follows:

$$\|\mathbf{u}\|_p = \left(\sum_i |u_i|^p \right)^{1/p} \quad \text{if } p \geq 1 . \quad (3)$$

The L_0 norm ($p = 0$) will be used extensively in this paper:

$$\|\mathbf{u}\|_0 = \sum_i \text{Ind}(u_i) , \quad (4)$$

where $\text{Ind}(\bullet)$ is a binary indicator function:

$$\text{Ind}(u_i) = \begin{cases} 1 & \text{if } u_i \neq 0 \\ 0 & \text{if } u_i = 0 \end{cases} . \quad (5)$$

Generally speaking, the L_0 norm measures the sparseness of a vector by counting the number of non-zero elements in the vector. However, it is worthwhile to mention that the “ L_0 norm” is, despite the terminology, not formally a norm since the basic property of *homogeneity of norm* is generally not satisfied, *i.e.*, $\|a\mathbf{u}\|_0 \neq |a| \|\mathbf{u}\|_0$.

In practice, the dimension of \mathbf{J} is generally larger than the number of available DFT reference calculations, *i.e.*, the equation system is underdetermined, so that Eq. (2) possesses an infinite number of solution vectors with an RMSE of 0. The direct solution of Eq. (2) therefore results in *over-fitting*, where the in-sample data is

perfectly reproduced at the cost of losing predictive power on out-of-sample data. A typical result of over-fitting in CE is that the ECI vector, \mathbf{J} , has many non-zero elements of large magnitude but with opposite signs. To avoid over-fitting, compressive sensing [19,28] adds L_1 norm regularization to the fitting procedure, resulting in the objective function

$$\min_{\mathbf{J}} \left\| \mathbf{E}_{\text{DFT,S}} - \mathbf{\Pi}_s \mathbf{J} \right\|_2^2 + \mu_1 \|\mathbf{J}\|_1 \quad (6)$$

where $\mu_1 > 0$ is the L_1 norm regularization constant. Candès *et al.* proved that minimizing the L_1 norm is, for many problems, equivalent to minimizing the L_0 norm [25], so that the L_1 norm regularization term $\mu_1 \|\mathbf{J}\|_1$ is essentially a penalty that forces \mathbf{J} to be sparse with components of relatively small magnitude, reducing over-fitting and allowing the construction of a robust CE model. However, the proof of reference [25] requires that the correlation matrix $\mathbf{\Pi}_s$ satisfies the restricted isometry property (RIP) [25] for Eq. (6) to provide the most accurate and sparse solution. In essence, RIP requires $\mathbf{\Pi}_s$ to be close to orthonormal (every principle submatrix of $\mathbf{\Pi}_s^T \mathbf{\Pi}_s$ to be close to identity matrix).

As discussed in reference [19], to satisfy RIP in the context of CE models, the reference lattice configurations in $\mathbf{\Pi}_s$ could be drawn such that the correlations cover the correlation space randomly. However, many relevant physical properties, such as thermodynamic phase diagrams, depend on highly accurate low-energy states but tolerate larger errors for lattice configurations with high energies. From that perspective, $\mathbf{\Pi}_s$ should preferably contain many more low-energy states than high energy states as it is more important that the CE model is able to distinguish between low energy states. Unfortunately, such a non-randomly sampled $\mathbf{\Pi}_s$ does not generally satisfy the RIP, and solving Eq. (6) directly may not lead to optimal results.

L_0L_1 normalization

The goal of this paper is, thus, to develop a CE construction based on a compressive sensing paradigm without the RIP assumption. The original compressive sensing CE paradigm aims at using L_1 norm regularization to reproduce the results of L_0 norm regularization, as L_1 norm regularization can be solved computationally efficiently using quadratic programming techniques [26]. Our proposed approach is to reintroduce L_0 norm regularization as the exact measure of sparseness into Eq. (6) to obtain a sparser and more accurate solution. In other words, the objective function becomes

$$\min_{\mathbf{J}} \|\mathbf{E}_{\text{DFT,S}} - \mathbf{\Pi}_s \mathbf{J}\|_2^2 + \mu_1 \|\mathbf{J}\|_1 + \mu_0 \|\mathbf{J}\|_0 \quad (7)$$

with $\mu_0 > 0$ being the L_0 norm regularization constant. The L_1 norm term in Eq. (7) is retained to assist the L_0 norm regularization, as the solution of L_1 norm minimization is already a close approximate to the exact solution of Eq. (7). Thus, intuitively speaking, the L_1 norm term guides the search for a better L_0 norm regularization result. The downside of this construction is that the L_0 norm minimization problem in Eq. (7) is NP-hard [29]. Practically, this means no algorithm can guarantee to obtain its exact solution in polynomial time. However, our main goal is to improve the CE model construction results beyond the conventional L_1 norm regularization, and any improved solution to Eq. (7) is desirable as long as it can be obtained within practical time limits.

Mixed integer quadratic programming for L_0L_1 -norm compressive sensing

To implement CE model construction based on the objective function of Eq. (7), we make use of a mathematical programming [30] technique called *mixed integer quadratic programming* (MIQP) [31,32]. In recent years, important developments have been made in heuristic methods [33] for the solution of MIQP without proof of optimality, including 1-opt, 2-opt [34], local branching [35], Relaxation Induced

Neighborhood Search (RINS) [36] and polishing [37] based on evolutionary algorithms. These heuristics enable the practical and efficient search for near-optimal solutions even if the underlying problem is NP-hard. Exact methods (with proof of optimality) including branch-and-bound [38,39], cutting planes [40,41] and branch-and-cut [42,43] have also been developed extensively. Building on these theoretical developments, there have been immense efforts to develop robust and efficient state-of-the-art solvers such as Gurobi [44] that fine-tune and integrate different heuristics and exact methods.

To solve the L_0L_1 -norm compressive sensing problem of Eq. (7), we convert Eq. (7) into an MIQP problem and use Gurobi to solve for its near optimal solution within a practical time span. The precise definition of MIQP and the conversion of Eq. (7) into an MIQP problem are elaborated in the following.

MIQP is a mathematical optimization technique for problems of the standard form:

$$\begin{aligned}
& \min_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \\
& \text{s.t. } \mathbf{A} \mathbf{x} \leq \mathbf{b} \\
& \quad \mathbf{C} \mathbf{x} = \mathbf{d} \\
& \quad x_i \in \mathbb{Z} \quad \forall i \in I
\end{aligned} \tag{8}$$

where \mathbf{Q} is a positive semi-definite matrix, i.e., for all real vectors \mathbf{x} , $\mathbf{x}^T \mathbf{Q} \mathbf{x} \geq 0$. The positive semi-definite property ensures that the function to be optimized is convex. \mathbf{A} and \mathbf{C} are real matrices, and \mathbf{b} , \mathbf{c} , and \mathbf{d} are real vectors. \mathbb{Z} is the set of integers and I is a set of indices for \mathbf{x} to be integers. Note that replacing \mathbb{Z} with $\{0,1\}$ leads to an important subclass of MIQP [45], due to the relationship:

$$x_i \in \{0,1\} \Leftrightarrow x_i \in \mathbb{Z} \text{ and } 0 \leq x_i \leq 1. \tag{9}$$

We convert Eq. (7) to an MIQP by expanding the quadratic term:

$$\min_{\mathbf{J}} \mathbf{J}^T \Pi_S^T \Pi_S \mathbf{J} - 2 \mathbf{E}_{\text{DFT},S}^T \Pi_S \mathbf{J} + \mathbf{E}_{\text{DFT},S}^T \mathbf{E}_{\text{DFT},S} + \mu_1 \|\mathbf{J}\|_1 + \mu_0 \|\mathbf{J}\|_0. \tag{10}$$

Note that the term $\mathbf{E}_{\text{DFT},S}^T \mathbf{E}_{\text{DFT},S}$ just gives rise to a constant shift that has no effect on the optimal solution \mathbf{J} . $\Pi_S^T \Pi_S$ is positive semi-definite because $\mathbf{x}^T \Pi_S^T \Pi_S \mathbf{x} = \|\Pi_S \mathbf{x}\|_2^2 \geq 0$. The key step to further convert Eq. (10) into a MIQP is to introduce auxiliary variables $z_{0,c}$ and $z_{1,c}$ for each cluster $c \in \mathbf{C}$ to track $\text{Ind}(J_c)$ and $|J_c|$ respectively. Explicitly, Eq. (10) is equivalent to:

$$\begin{aligned}
& \min_{\mathbf{J}, \mathbf{z}} \mathbf{J}^T \Pi_S^T \Pi_S \mathbf{J} - 2\mathbf{E}_{\text{DFT},S}^T \Pi_S \mathbf{J} + \mathbf{E}_{\text{DFT},S}^T \mathbf{E}_{\text{DFT},S} + \mu_1 \sum_{c \in \mathbf{C}} z_{1,c} + \mu_0 \sum_{c \in \mathbf{C}} z_{0,c} \\
& \text{st. } z_{1,c} \geq J_c \quad \forall c \in \mathbf{C} \\
& \quad z_{1,c} \geq -J_c \quad \forall c \in \mathbf{C} \\
& \quad Mz_{0,c} \geq J_c \quad \forall c \in \mathbf{C} \\
& \quad Mz_{0,c} \geq -J_c \quad \forall c \in \mathbf{C} \\
& \quad z_{0,c} \in \{0,1\} \quad \forall c \in \mathbf{C}
\end{aligned} \tag{11}$$

where M is some large number which $|J_c|$ should never exceed for every cluster $c \in \mathbf{C}$. For example, M could be set to 50 *eV/f.u.* (electronvolt per formula unit) as long as we know intuitively that the magnitude of ECIs should never exceed 50 *eV/f.u.* Inclusion of such M parameters is generally referred to as “big M formulation” in mathematical programming [30].

A proof of the equivalence between Eq. (10) and Eq. (11) is provided in the following section.

Proof of the equivalence of the L_0L_1 -norm regularized cluster expansion objective function Eq. (10) and its MIQP reformulation Eq. (11)

To prove the equivalence of Eq. (10) and Eq. (11), we show:

(A) For any optimal solution $(\mathbf{J}', \mathbf{z}', \mathbf{z}'_0)$ to Eq. (11), $\mathbf{J} = \mathbf{J}'$ is an optimal solution to Eq. (10).

(B) For any optimal solution \mathbf{J} to Eq. (10), $\mathbf{J}' = \mathbf{J}$, $z'_{1,c} = |J_c|$ and $z'_{0,c} = \text{Ind}(J_c)$, for all $c \in \mathbf{C}$ is an optimal solution to Eq. (11).

An important lemma (L1) needs to be established before the proof of (A), (B).

(L1) $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ is optimal to Eq. (11) implies $z'_{1,c} = |J'_c|$ and $z'_{0,c} = \text{Ind}(J'_c) \quad \forall c \in \mathbf{C}$.

Proof of (L1):

(i) $z'_{1,c} = |J'_c| \quad \forall c \in \mathbf{C}$. If $z'_{1,c} < |J'_c|$, it violates either $z'_{1,c} \geq J'_c$ or $z'_{1,c} \geq -J'_c$. If $z'_{1,c} > |J'_c|$, the objective function in Eq. (11) is strictly greater than for $z'_{1,c} = |J'_c|$ since $\mu_1 > 0$. Therefore $z'_{1,c} = |J'_c|$.

(ii) $z'_{0,c} = \text{Ind}(J'_c) \quad \forall c \in \mathbf{C}$. Note that $z'_{0,c} \in \{0,1\}$ and $\text{Ind}(J'_c) \in \{0,1\}$. Suppose $z'_{0,c} = 0$ and $\text{Ind}(J'_c) = 1$. We know that $J'_c \neq 0$ (since $\text{Ind}(J'_c) = 1$). However, the required constraints $Mz'_{0,c} = 0 \geq J'_c$ and $-Mz'_{0,c} = 0 \leq J'_c$ would lead to $J'_c = 0$. This is a contradiction and thus $\text{Ind}(J'_c) = 1$ implies $z'_{0,c} = 1$. Suppose $z'_{0,c} = 1$ and $\text{Ind}(J'_c) = 0$. We know that $J'_c = 0$, since $\text{Ind}(J'_c) = 0$. Taking $z'_{0,c} = 0$ leads to a strictly lower objective function value in Eq. (11) and no constraints are violated. This also leads to a contradiction since some parameters lead to a lower objective function value than $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$. Therefore, $\text{Ind}(J'_c) = 0$ implies $z'_{0,c} = 0$. We thus conclude that $z'_{0,c} = \text{Ind}(J'_c)$. Therefore, if $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ is optimal for Eq. (11), this implies that $z'_{1,c} = |J'_c|$ and $z'_{0,c} = \text{Ind}(J'_c) \quad \forall c \in \mathbf{C}$.

Proof of (A):

Suppose $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ is optimal to Eq. (11) but $\mathbf{J} = \mathbf{J}'$ is not optimal to Eq. (10). Then there must be some other \mathbf{J}'' that optimizes Eq. (10). This means Eq. (10) evaluated at \mathbf{J}'' is strictly less than Eq. (10) evaluated at \mathbf{J}' . From (L1), we know that $z'_{1,c} = |J'_c|$ and $z'_{0,c} = \text{Ind}(J'_c) \quad \forall c \in \mathbf{C}$. Therefore, Eq. (11) evaluated at

$\left(\mathbf{J}', (|J'_c|)_{c \in \mathbf{C}}, (\text{Ind}(J'_c))_{c \in \mathbf{C}}\right)$ is equal to Eq. (10) evaluated at \mathbf{J}' . Note that $\left(\mathbf{J}'', (|J''_c|)_{c \in \mathbf{C}}, (\text{Ind}(J''_c))_{c \in \mathbf{C}}\right)$ is feasible (satisfying all constraints) to Eq. (11), and Eq. (11) evaluated at $\left(\mathbf{J}'', (|J''_c|)_{c \in \mathbf{C}}, (\text{Ind}(J''_c))_{c \in \mathbf{C}}\right)$ is equal to Eq. (10) evaluated at \mathbf{J}'' . This implies that Eq. (11) evaluated at $\left(\mathbf{J}'', (|J''_c|)_{c \in \mathbf{C}}, (\text{Ind}(J''_c))_{c \in \mathbf{C}}\right)$ is strictly less than Eq. (11) evaluated at $\left(\mathbf{J}', (|J'_c|)_{c \in \mathbf{C}}, (\text{Ind}(J'_c))_{c \in \mathbf{C}}\right)$, contradicting the assumption that $\left(\mathbf{J}', (|J'_c|)_{c \in \mathbf{C}}, (\text{Ind}(J'_c))_{c \in \mathbf{C}}\right)$ is optimal. Therefore, if $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ is optimal for Eq. (11), this implies that \mathbf{J}' is optimal for Eq. (10).

Proof of (B):

Suppose \mathbf{J} optimizes Eq. (10). Consider any $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ that is optimal for Eq. (11). Using (A), we know that \mathbf{J}' is then also optimal for Eq. (10). Therefore, Eq. (10) evaluated at \mathbf{J} is equal to Eq. (10) evaluated at \mathbf{J}' . Using (L1) we know that $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0) = \left(\mathbf{J}', (|J'_c|)_{c \in \mathbf{C}}, (\text{Ind}(J'_c))_{c \in \mathbf{C}}\right)$. Therefore Eq. (11) evaluated at $(\mathbf{J}', \mathbf{z}'_1, \mathbf{z}'_0)$ is equal to Eq. (10) evaluated at \mathbf{J}' . Note that Eq. (11) evaluated at $\left(\mathbf{J}, (|J_c|)_{c \in \mathbf{C}}, (\text{Ind}(J_c))_{c \in \mathbf{C}}\right)$ is equal to Eq. (10) evaluated at \mathbf{J} . Therefore, $\left(\mathbf{J}, (|J_c|)_{c \in \mathbf{C}}, (\text{Ind}(J_c))_{c \in \mathbf{C}}\right)$ is also an optimal solution to Eq. (11).

The equivalence between Eq. (10) and Eq. (11) is proved as a consequence of the proofs of (A) and (B). One consequence of this equivalence is that the constraints in Eq. (11) automatically ensure $z'_{1,c} = |J'_c|$ and $z'_{0,c} = \text{Ind}(J'_c)$, and therefore the solution spaces of Eq. (10) and Eq. (11) are equivalent.

Comparison of the L_0L_1 -norm formulation with the L_1 -norm formulation

In this section, we provide a mathematical explanation of why solutions to the L_0L_1 -norm formulation may be closer to those of the ideal L_0 -norm formulation than solutions to the L_1 -norm formulation. Additionally, we also show that if the L_1 -norm formulation provides an exact solution to the L_0 -norm formulation (which is the case under certain conditions), then the L_0L_1 -norm formulation also provides an exact solution to the L_0 -norm formulation (although the solutions may be associated with different parameters).

From an intuitive perspective, the L_0L_1 -norm formulation simply biases the solution more towards the L_0 norm, and so intuitively the solution associated with the L_0L_1 -norm formulation is closer to that of L_0 norm for an equivalent L_2 -norm error. This relationship can also be shown with formal arguments:

Let J_1^* be an optimal solution to the L_1 -norm formulation of Eq. (6) with regularization parameter μ_1 . There exists ε_1 such that J_1^* is an optimal solution to

$$\begin{aligned} \min \quad & \|\mathbf{J}\|_1 \\ \text{s.t.} \quad & \|\mathbf{E}_{\text{DFT},S} - \mathbf{\Pi}_S \mathbf{J}\|_2^2 \leq \varepsilon_1 \end{aligned} \quad (12)$$

Now let J_{10}^* be an optimal solution to the L_0L_1 -norm formulation, Eq. (7), with regularization parameter μ'_0 and μ'_1 . There exist ε_{10} such that J_{10}^* is an optimal solution to

$$\begin{aligned} \min \quad & \|\mathbf{J}\|_1 + \frac{\mu'_0}{\mu'_1} \|\mathbf{J}\|_0 \\ \text{s.t.} \quad & \|\mathbf{E}_{\text{DFT},S} - \mathbf{\Pi}_S \mathbf{J}\|_2^2 \leq \varepsilon_{10} \end{aligned} \quad (13)$$

In order to compare the proximity of the L_0L_1 -norm formulation to the L_0 -norm formulation, we need to have the same reference. Therefore, without loss of generality, we choose μ_1 , μ'_0 and μ'_1 such that

$$\varepsilon_1 = \varepsilon_{10} \equiv \varepsilon_0 \quad (14)$$

where ε_0 is defined as the common constraint parameter. We can now consider whether J_1^* or J_{10}^* is closer to the L_0 -norm solution. To achieve this, we consider the L_0 -norm problem

$$\begin{aligned} \min \quad & \|\mathbf{J}\|_0 \\ \text{s.t.} \quad & \|\mathbf{E}_{\text{DFT},S} - \mathbf{\Pi}_S \mathbf{J}\|_2^2 \leq \varepsilon_0 \end{aligned} \quad (15)$$

Note that J_1^* and J_{10}^* are both satisfying the constraint in Eq. (15) given (14).

We now show that $\|J_{10}^*\|_0 \leq \|J_1^*\|_0$ by contradiction:

Suppose $\|J_{10}^*\|_0 > \|J_1^*\|_0$. Since J_1^* is an optimal solution to Eq. (12), we conclude

$\|J_{10}^*\|_1 \geq \|J_1^*\|_1$. Therefore, $\|J_{10}^*\|_1 + \frac{\mu'_0}{\mu'_1} \|J_{10}^*\|_0 > \|J_1^*\|_1 + \frac{\mu'_0}{\mu'_1} \|J_1^*\|_0$, contradicting that J_{10}^* is an

optimal solution to Eq. (13). Hence, we have shown that

$$\|J_{10}^*\|_0 \leq \|J_1^*\|_0 \quad (16)$$

We conclude that given the same L_2 -norm error, the L_0L_1 -norm formulation provides a solution that is at least as good as the solution to the L_1 -norm formulation assuming that the L_0 -norm formulation is the targeted solution. Therefore, when R.I.P. is satisfied and the L_1 -norm formulation provides the exact solution to the L_0 -norm problem, the L_0L_1 -norm formulation also provides an exact solution as the L_0 -norm formulation given the same L_2 -norm error.

Hierarchical constraints

One advantage of the general L_0L_1 formulation of Eq.(11) is that it allows introducing additional constraints in a straightforward fashion. Here we demonstrate this concept by including an intuitive hierarchy constraint on the cluster basis. Similar concepts have previously been proposed in the literature [21,22].

One basic assumption of CE models is that n -body contributions to the configuration energy become less important the larger n becomes, so that the expansion can be truncated at a certain n . Often, 4-body interactions, *i.e.*, quadruplet clusters, are taken as the limit. To incorporate this principle in the model construction and to further reduce over-fitting, we introduce one more type of constraint into Eq. (11). This constraint is motivated by the natural hierarchy within the set of clusters \mathbf{C} . An example of this hierarchy is that a triplet is composed of three pairs (some of which may be equivalent due to symmetry), so the triplet has a higher order than any of these three pairs. If we consider some pair that is not contained within the triplet, we say that the pair and the triplet are incomparable. More generally, we define cluster a to have higher order than cluster b if cluster a contains cluster b . We further define the order of cluster a to be exactly one higher than the order of cluster b , if cluster a has higher order than cluster b and the size of cluster a is equal to the size of cluster b plus 1. The hierarchical set H is defined as

$$H = \{(a,b) \in \mathbf{C} \times \mathbf{C} : \text{cluster } a \text{ is one order higher than cluster } b\}. \quad (17)$$

Generally speaking, a higher order cluster can be thought of as a correction to lower order clusters when the lower order clusters are not sufficient to describe all interactions. To incorporate this physical intuition, we add the following type of constraints into Eq. (11):

$$z_{0,a} \leq z_{0,b} \quad \forall (a,b) \in H. \quad (18)$$

Now, if $J_a \neq 0$, the original constraints in Eq. (11) lead to $z_{0,a} = 1$, and the additional constraint in Eq. (18) leads to $z_{0,b} = 1$.

We define any cluster c to be active if $z_{0,c} = 1$. Note that when the constraint of Eq. (18) is integrated into Eq. (11), a lower order cluster b automatically becomes active when a related higher order cluster a is active. Conversely, b being active no longer implies $J_b \neq 0$, but $J_b \neq 0$ does imply that b is active. In other words, when a higher order cluster a needs to be active, Eq. (18) requires that every sub-cluster b

of cluster a becomes also active, resulting in greater penalties to activate higher order clusters. As there are more higher order clusters than lower order clusters, Eq. (18) essentially guides the search of necessary higher order clusters in the L_0L_1 -norm compressive sensing paradigm.

An illustration of the hierarchy constraint outlined above is shown in Figure 1 for a simple example lattice with four sites. The largest cluster in the example is a quadruplet cluster. When the ECI of that cluster becomes greater than zero, *i.e.*, when the quadruplet cluster becomes active, all smaller clusters also become active. This means, the L_0 regularization parameters of all lower-order clusters c become $z_{0,c} = 1$, owing to the constraint Eq. (18).

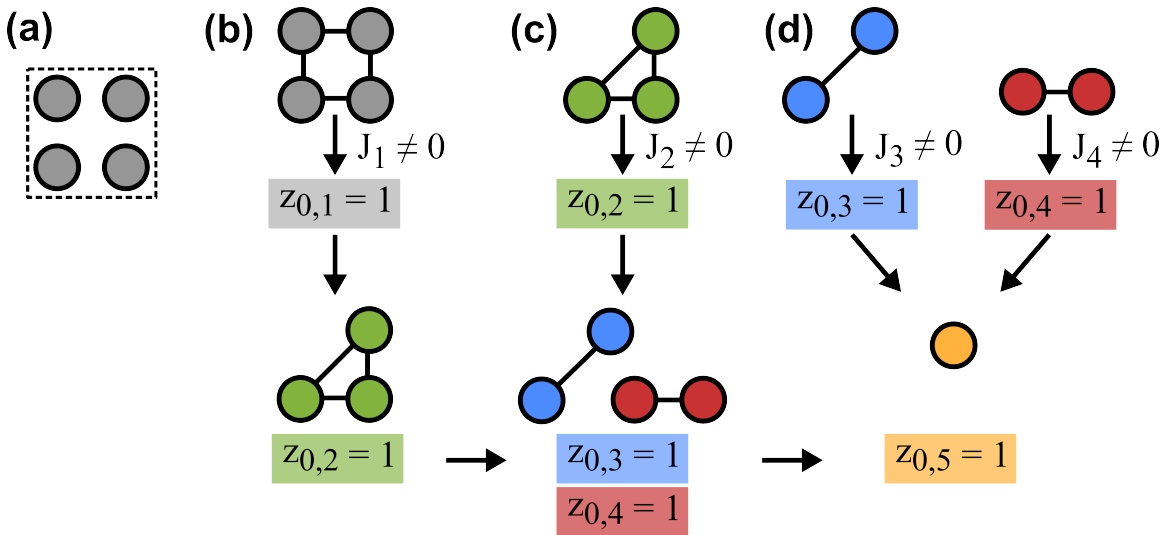


Figure 1, cluster hierarchy constraint for an example lattice. (a) The example lattice with four equivalent sites. (b) If the ECI of the quadruplet cluster is different from zero, all smaller (lower-order) clusters become automatically active. (c) Activating the triplet cluster does not affect the quadruplet cluster as it is of higher order, but it results in activation of both pair clusters and the point cluster. (d) Activating any of the two pair clusters will also activate the point cluster.

Results

Applications to cathode transition-metal oxide systems

Intercalation batteries function by reversible extraction and reintercalation of alkali metal or alkaline metal working ions from the cathode material, which typically is a transition-metal (TM) oxide or phosphate [46]. The static ordering of the working ions (*e.g.*, Li^+ , Na^+ , Mg^{2+}) and the TM cations within the crystal structure of the cathode affect the achievable capacity and rate capability. Additionally, when working ions are extracted from the cathode, additional ordering interaction between the remaining working ions and vacancies can strongly affect the voltage profile and have to be considered in the computational design of novel battery materials. Owing to the large number of configurational degrees of freedom, determining the most stable working-ion/TM/vacancy orderings in realistic cathode materials at different states of charge is usually not feasible with DFT calculations and is a prototypical application of CE model simulations [47].

To benchmark our L_0L_1 -norm compressive sensing CE approach, we apply the method to ordering problems in four different battery cathode systems: binary Li/Ti ordering in rocksalt-type $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ [48,49], Mg/vacancy ordering in spinel-type $\text{Mg}_x\text{Cr}_2\text{O}_4$ [50], Na/vacancy O3-type layered Na_xCrO_2 [51], and ternary Li/Ni/vacancy ordering in rocksalt- and spinel-type $\text{Li}_x\text{Ni}_y\text{O}$ [52]. The different types of configurational ordering problems in the chosen materials provide a comprehensive benchmark of the new methodology.

All reference DFT energies of configurations used in the CE model construction were calculated under the Generalized Gradient Approximation with the Perdew-Burke-Ernzerhof (PBE) functional [53]. Further, a Hubbard-U correction [54,55] was employed for the d states of Ni and Cr using the U values from the work of Jain et al [56]. A plane wave basis set with an energy cutoff of 520 eV and projector-augmented wave (PAW) pseudopotentials were employed to describe the electronic wavefunctions [57,58]. All DFT calculations were done with the Vienna Ab initio

Simulation Package (VASP) [59,60]. For the chosen materials systems, initial sets of configurations were generated at various concentrations using an enumerating algorithm [61-63]. Subsequently, a ground-state searching algorithm for lattice models [16] was used to further expand the DFT dataset with additional low energy configurations. The geometry and cell parameters of all of the reference structures were optimized to nearby local minima. The final reference sets used for the CE model construction comprised 796 $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ configurations, 249 $\text{Mg}_x\text{Cr}_2\text{O}_4$ configurations, 281 Na_xCrO_2 configurations, and 453 $\text{Li}_x\text{Ni}_y\text{O}$ configurations.

To assess the effectiveness of our L_0L_1 -norm compressive sensing paradigm compared to the conventional L_1 -norm compressive sensing for practical CE applications, we consider the 10-fold cross validation score (cv score) and the sparseness (model complexity) as a dual metric. The cv score is defined as the root mean square of the out-of-sample error. The average sparseness is the average L_0 -norm of ECIs for each partition. In the case of the (conventional) L_1 -norm compressive sensing, cross validation is performed with μ_1 ranging from 0.001 to 10 in increments of 40 partitions on the logarithmic scale. For the L_0L_1 -norm compressive sensing, cross validation is performed with μ_1 ranging from 0.01 to 4 $eV/f.u.$, with 20 logarithmically-scaled increments and μ_0 taking on 6 different values: 3×10^{-5} , 1×10^{-4} , 3×10^{-4} , 1×10^{-3} , 3×10^{-3} , and 1×10^{-2} $eV/f.u.$. These ranges for μ_0 and μ_1 were empirically found to give the optimal results. Since the solution of the L_0L_1 -norm formulation is not guaranteed to terminate in a polynomial time, the MIQP solver is terminated after 300 seconds for each partition, and the best result from within this time frame is extracted. For each (μ_0, μ_1) pair, we record the average sparseness and the cv score.

To decide the current optimal CE fit, we define the *(sparseness, cv score)* front as follows: a *(sparseness, cv score)* pair is on the *(sparseness, cv score)* front if there exists no other *(sparseness', cv score')* such that $\text{sparseness}' \leq \text{sparseness}$ and $\text{cv score}' \leq \text{cv score}$. In other words, any *(sparseness, cv score)* on the *(sparseness, cv score)*

front is one optimal solution of the multi-objective minimization in the (*sparseness, cv score*) space. However, we emphasize that the standard choice of the final ECIs are those associated with the minimum *cv score*. Other ECIs are only chosen when there is a very strong preference for sparseness, a situation that has not occurred in our benchmark examples.

The (*sparseness, cv score*) front for the four different material systems is shown in Figure 2. We start our discussions with the comparison of the L_0L_1 -norm formulation (illustrated with the blue curve) with conventional L_1 -norm compressive sensing (illustrated with the red curve). For the $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ system (Figure 2(a)), the optimal (*sparseness, cv score*) obtained with the conventional L_1 -norm compressive sensing of Eq. (6) is $(129.7, 0.0346 \text{ eV/f.u.})$, while using the L_0L_1 -norm formulation of Eq. (11) and Eq. (18) achieves an optimal (*sparseness, cv score*) of $(38.9, 0.0334 \text{ eV/f.u.})$. Hence, our new approach leads to a CE model that is 70% more sparse with similar predictive error. Further, based on the sparseness front shown in

Figure 2(a), the L_0L_1 -norm formulation still outperforms the standard L_1 -norm compressive sensing when a higher *cv score* is traded in for a sparser CE. For a fixed *cv score*, the sparseness resulting from the L_0L_1 -norm formulation is generally around 70% sparser. Conversely, for the same sparseness, the *cv score* is on average around 30% lower under the L_0L_1 -norm compressive sensing paradigm. The above comparison indicates that the L_0L_1 -norm formulation has much better compression capability, *i.e.*, its ability to extract important features from the DFT reference data.

For $\text{Li}_x\text{Ni}_y\text{O}$, $\text{Mg}_x\text{Cr}_2\text{O}_4$, and Na_xCrO_2 , the results follow essentially the same trend with optimal CE models being 59%, 64%, and 66% sparser and 0.6%, 3.6%, and 0.7% more predictive in terms of *cv score*. The general trends and characteristics of the (*sparseness, cv score*) front are very similar for these materials, demonstrating the robustness and versatility of the L_0L_1 -norm formulation.

We proceed to discuss the effect of hierarchical constraints, Eq. (18), on the L_0L_1 -norm formulation. We compare the (*sparseness, cv score*) front for the L_0L_1 -norm formulation with (blue curve in Figure 2) and without (black curve in Figure 2) hierarchical constraints. As seen in the figure, the two formulations show similar behavior. Both L_0L_1 formulations show substantial and comparable improvement of the sparseness over the L_1 -norm compressive sensing of Eq. (6). For the $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ system (Figure 2 (a)), the optimal *cv score* obtained with just the L_0L_1 -norm formulation without hierarchical constraints is 0.0345 eV/f.u. , while using the L_0L_1 -norm formulation of Eq. (11) and Eq. (18) including the hierarchy constraints achieves an optimal *cv score* of 0.0334 eV/f.u. . Therefore, the hierarchical constraints have resulted in a *cv score* improvement of 3.3% for this system. For the $\text{Li}_x\text{Ni}_y\text{O}$ and Na_xCrO_2 systems, the optimal *cv scores* are similar. The inclusion of hierarchical constraints results in 0.4% and 0.5% *cv score* reduction respectively. However, for $\text{Mg}_x\text{Cr}_2\text{O}_4$ the inclusion of hierarchical constraints leads to an increased *cv score* (-1.7% worse performance). We emphasize that the benefit of the L_0L_1 -norm formulation is the capability to incorporate additional constraints, such as the physically intuitive hierarchy constraints, in a straightforward manner. This is however an entirely separate concern from the CE fit and it is well possible that intuitive constraints may lead to worse *cv scores* for specific systems though they might aid in the interpretation of the CE model. The remaining discussion in this paper assumes applications of the L_0L_1 -norm formulation including hierarchical constraints.

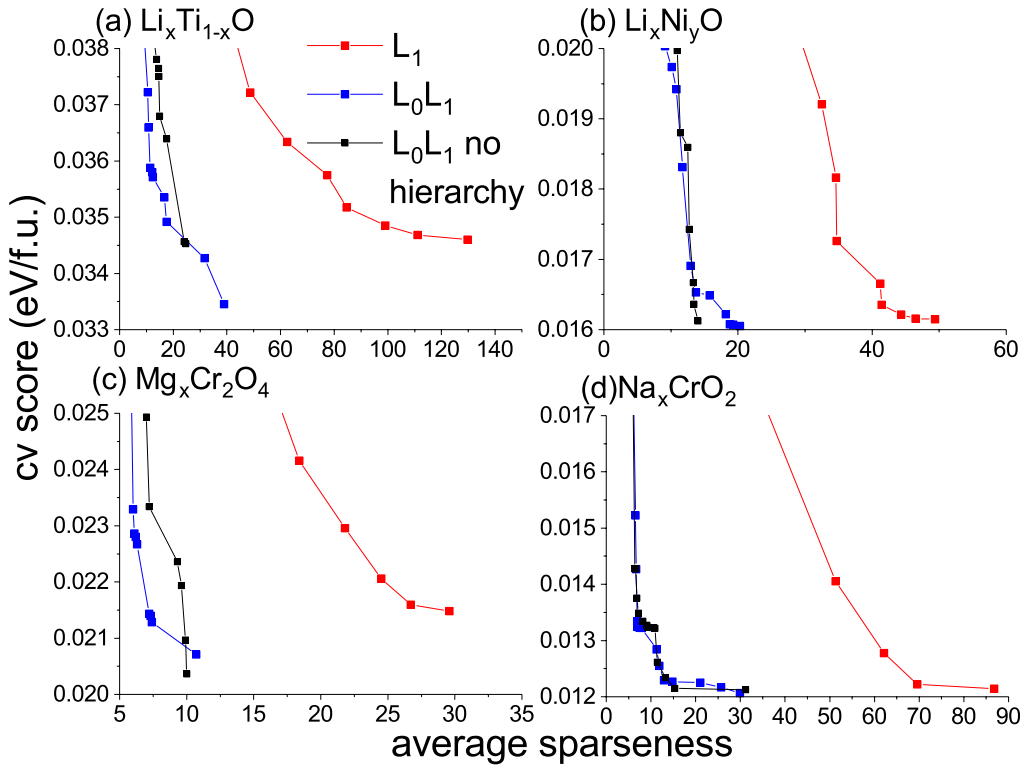


Figure 2, the (*sparseness, cv score*) front as a measure of the quality of the CE fit for four different materials: (a) $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$, (b) $\text{Li}_x\text{Ni}_y\text{O}$, (c) $\text{Mg}_x\text{Cr}_2\text{O}_4$, and (d) Na_xCrO_2 . The red line corresponds to the (*sparseness, cv score*) front constructed with standard L_1 -norm compressive sensing, whereas the blue line corresponds to that constructed with the L_0L_1 -norm paradigm and the black line corresponds to the L_0L_1 -norm paradigm without the hierarchical constraints.

To further analyze the sparseness improvements based on L_0L_1 -norm formulation, we examine the differences in the ECI sets in detail for the example of the $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ system. The ECIs resulting from both construction paradigms are plotted in Figure 3. The first eight ECIs (clusters 0 through 7) belong to the constant term, point term, and pair interactions. The triplet terms correspond to clusters 8 through 51, and the quadruplet terms are clusters 52 through 410. The vertical lines in Figure 3 indicate the limits of these three sets of ECIs. Clusters of the same order are sorted according to the maximal distance between two sites within them, which is an intuitive measure of the importance of the cluster.

Figure 3 (a) shows all ECIs for $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$. A first obvious difference between the two sets of ECIs is that there are only 7 nonzero ECIs beyond cluster 70 under the L_0L_1 -norm formulation, while there are 96 under L_1 -norm compressive sensing. Considering our initial assumption that the higher-order quadruplet terms are generally minor corrections to the pair and triplet interactions, one would expect that most of the corresponding ECIs should be 0. Clearly, from this physical perspective, the L_0L_1 -norm formulation conforms better to the sparse physics paradigm. Figure 3 (b) shows only the ECIs up to cluster 70, which comprise all constant, point, pair, and triplet terms as well as the most important quadruplets with short internal distances. Interestingly, for the constant, point, and pair terms, the two sets of ECIs are very similar. The same is true for short distance triplets and quadruplets for which the L_0L_1 -norm and L_1 -norm formulations also generally agree with each other, indicating that both formulations capture the important short-range interactions well.

The key advantage of the L_0L_1 -norm formulation is that higher-order clusters that correspond to longer ranged interactions are more compressed, in accordance with the expectations of the sparse physics paradigm. CE models with fewer clusters can be evaluated more efficiently, so that the sparser CE models result in a direct improvement of the computational efficiency. Additionally, the interpretation of the physics of a material is more straightforward for models with low complexity.

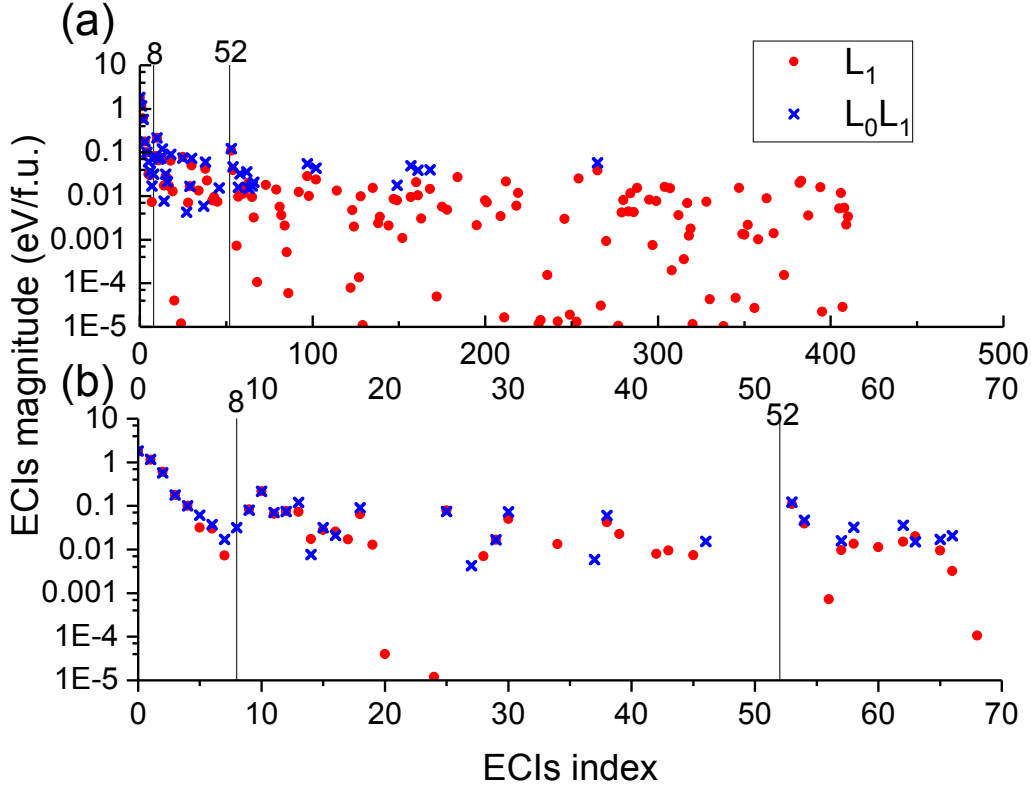


Figure 3 Comparison of ECIs for $\text{Li}_x\text{Ti}_{(1-x)}\text{O}$ obtained from conventional L_1 -norm compressive sensing (red circles) and from our proposed L_0L_1 -norm formulation (blue crosses). The vertical lines at index 8 and 52 indicate the transition from pairs to triplets and from triplets to quadruplets, respectively. (a) shows all of the ECIs, (b) shows only the ECIs up to index 70.

In summary, we have derived an L_0L_1 -norm compressive sensing paradigm for the construction of cluster expansion lattice models from reference configurations that are not required to possess the restricted isometry property. We present an efficient implementation of this new paradigm using state-of-the-art mixed-integer quadratic programming techniques. We demonstrated the effectiveness and robustness of the method for four different transition-metal oxides with relevance as battery cathodes. The L_0L_1 -norm compressive sensing formulation consistently results in significantly sparser CE models with predictive power as good or better than CE models constructed using conventional L_1 -norm compressive sensing.

Competing interests

The authors declare no competing financial interests.

Acknowledgements

This work was primarily funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05-CH11231 (Materials Project program KC23MP). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant no. ACI-1053575, and resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the US Department of Energy under contract no. DE-C02-05CH11231.

References

- [1] P. Hohenberg and W. Kohn, Phys Rev B **136**, B864 (1964).
- [2] W. Kohn and L. J. Sham, Phys Rev **140**, 1133 (1965).
- [3] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, Nat Mater **5**, 641 (2006).
- [4] K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, and S. Curtarolo, Nat Mater **11**, 614 (2012).
- [5] J. K. Norskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nat Chem **1**, 37 (2009).
- [6] Z. Rong, R. Malik, P. Canepa, G. Sai Gautam, M. Liu, A. Jain, K. Persson, and G. Ceder, Chemistry of Materials **27**, 6016 (2015).
- [7] J. M. Seminario, *Recent developments and applications of modern density functional theory* (Elsevier, 1996), Vol. 4.
- [8] I. V. Solovyev, J Phys-Condens Mat **20**, 293201 (2008).
- [9] M. Liu, Z. Rong, R. Malik, P. Canepa, A. Jain, G. Ceder, and K. A. Persson, Energy & Environmental Science **8**, 964 (2015).
- [10] A. Jain, Y. Shin, and K. A. Persson, Nature Reviews Materials **1**, 15004 (2016).
- [11] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A: Statistical Mechanics and its Applications **128**, 334 (1984).
- [12] A. van de Walle and G. Ceder, J Phase Equilib **23**, 348 (2002).

- [13] D. De Fontaine, Solid state physics **34**, 73 (1979).
- [14] D. De Fontaine, Solid state physics **47**, 33 (1994).
- [15] W. Huang, D. Kitchaev, S. Dacek, Z. Rong, Z. Ding, and G. Ceder, arXiv preprint arXiv:1606.07429 (2016).
- [16] W. Huang, D. A. Kitchaev, S. T. Dacek, Z. Rong, A. Urban, S. Cao, C. Luo, and G. Ceder, Phys Rev B **94**, 134424 (2016).
- [17] L. M. Herder, J. M. Bray, and W. F. Schneider, Surf Sci **640**, 104 (2015).
- [18] V. Blum, G. L. Hart, M. J. Walorski, and A. Zunger, Phys Rev B **72**, 165113 (2005).
- [19] L. J. Nelson, G. L. W. Hart, F. Zhou, and V. Ozoliņš, Phys Rev B **87**, 035125 (2013).
- [20] L. J. Nelson, V. Ozolins, C. S. Reese, F. Zhou, and G. L. W. Hart, Phys Rev B **88**, 155105 (2013).
- [21] N. A. Zarkevich and D. D. Johnson, Physical review letters **92**, 255702 (2004).
- [22] M. H. Sluiter and Y. Kawazoe, Phys Rev B **71**, 212201 (2005).
- [23] E. Cockayne and A. van de Walle, Phys Rev B **81**, 012104 (2010).
- [24] E. J. Candès, J. Romberg, and T. Tao, IEEE Transactions on information theory **52**, 489 (2006).
- [25] E. J. Candès, Comptes Rendus Mathematique **346**, 589 (2008).
- [26] W. Huang, A. Urban, Z. Rong, Z. Ding, C. Luo, and G. Ceder, npj Computational Materials **3**, 30 (2017).
- [27] G. Ceder, Computational Materials Science **1**, 144 (1993).
- [28] E. J. Candès and M. B. Wakin, IEEE signal processing magazine **25**, 21 (2008).
- [29] D. Ge, X. Jiang, and Y. Ye, Math Program **129**, 285 (2011).
- [30] W. L. Winston, M. Venkataramanan, and J. B. Goldberg, *Introduction to mathematical programming* (Thomson/Brooks/Cole Duxbury; Pacific Grove, CA, 2003), Vol. 1.
- [31] R. Lazimy, Math Program **22**, 332 (1982).
- [32] P. Bonami, M. Kiliç, and J. Linderoth, in *Mixed integer nonlinear programming* (Springer, 2012), pp. 1.
- [33] T. Achterberg and R. Wunderling, in *Facets of Combinatorial Optimization* (Springer, 2013), pp. 449.
- [34] G. A. Croes, Operations research **6**, 791 (1958).
- [35] M. Fischetti and A. Lodi, Math Program **98**, 23 (2003).
- [36] E. Danna, E. Rothberg, and C. L. Pape, Math Program **102**, 71 (2005).
- [37] E. Rothberg, INFORMS Journal on Computing **19**, 534 (2007).
- [38] A. H. Land and A. G. Doig, 50 Years of Integer Programming 1958-2008, 105 (2010).
- [39] W. Cook, Optimization Stories, 227 (2012).
- [40] R. Gouonr, Bull. Am. Math. Soc **64**, 3 (1958).
- [41] G. B. Dantzig, D. R. Fulkerson, and S. M. Johnson, Operations Research **7**, 58 (1959).
- [42] M. Padberg and G. Rinaldi, Operations Research Letters **6**, 1 (1987).
- [43] M. Padberg and G. Rinaldi, SIAM review **33**, 60 (1991).
- [44] G. Optimization, URL: <http://www.gurobi.com> **2**, 1.3 (2012).
- [45] J. E. Beasley, London, UK: Management School, Imperial College **4** (1998).

- [46] M. S. Whittingham, *Chemical reviews* **104**, 4271 (2004).
- [47] A. J. Toumar, S. P. Ong, W. D. Richards, S. Dacek, and G. Ceder, *Physical Review Applied* **4**, 064002 (2015).
- [48] P. Bongers, University of Leiden thesis (1957).
- [49] A. Lecerf, in *ANNALES DE CHIMIE FRANCE*1962), pp. 513.
- [50] H. S. C. O'Neill and W. Dollase, *Physics and Chemistry of Minerals* **20**, 541 (1994).
- [51] S. Komaba, T. Nakayama, A. Ogata, T. Shimizu, C. Takei, S. Takada, A. Hokura, and I. Nakai, *Ecs Transactions* **16**, 43 (2009).
- [52] J. N. Reimers, J. Dahn, J. Greedan, C. Stager, G. Liu, I. Davidson, and U. Von Sacken, *Journal of Solid State Chemistry* **102**, 542 (1993).
- [53] J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical review letters* **77**, 3865 (1996).
- [54] V. V. Anisimov, J. Zaanen, and O. K. Andersen, *Phys Rev B Condens Matter* **44**, 943 (1991).
- [55] S. Dudarev, G. Botton, S. Savrasov, C. Humphreys, and A. Sutton, *Phys Rev B* **57**, 1505 (1998).
- [56] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, *Computational Materials Science* **50**, 2295 (2011).
- [57] P. E. Blöchl, *Phys Rev B* **50**, 17953 (1994).
- [58] G. Kresse and D. Joubert, *Phys Rev B* **59**, 1758 (1999).
- [59] G. Kresse and J. Furthmüller, *Phys Rev B* **54**, 11169 (1996).
- [60] G. Kresse and J. Furthmüller, *Computational materials science* **6**, 15 (1996).
- [61] G. L. Hart and R. W. Forcade, *Phys Rev B* **77**, 224115 (2008).
- [62] G. L. Hart and R. W. Forcade, *Phys Rev B* **80**, 014120 (2009).
- [63] G. L. Hart, L. J. Nelson, and R. W. Forcade, *Computational Materials Science* **59**, 101 (2012).