review from referee 1

Journal: Acta Crystallographica Section A: Foundations and Advances
Paper: uv5018
Authors: Lawrence C. Andrews*, Herbert J. Bernstein and Nicholas K. Sauter
Title: Approximating Lattice Similarity

Addendum to my previous report:

For clarity, I recommend the following parts to be omitted from the manuscript:

2.2 The space G6 (as it is not directly being used in the work)

Table 9 (as it doesn't contain the intermediate points and thus already being included in Table 10)

****
The github link incorrectly contains the final dot in the link, i.e., the link should be "https://github.com/duck10/LatticeRepLib.git" as opposed to "https://github.com/duck10/LatticeRepLib.git." (and please accept my personal gratitude for supporting free & open software!)

review from referee 1

Referee 1 attached the following notes to their review.

The authors have explored an application of a concept that was also introduced by themselves for the comparison of lattices by transforming them to a point that is in the proximity of a reference lattice's corresponding point in S6 space and have demonstrated the application of their method over a couple of case studies.

With all due respect, below are my comments, suggestions and questions. The suggestions for grammar and syntax should be taken as mere suggestions but I would be happy if my questions were to be corresponded.

[p2#33] G6 space is not used and actually being left out in favor of the S6 space [p3#34], therefore making its definition and mention redundant.

[p2#46] Referring an entry from Wikipedia for the definition of "similarity" is not recommended as Wikipedia is prone to errors and misleading information compared to academical texts. As a suggestion, the reference in the related Wikipedia entry ("Geometry and Symmetry" by Paul B. Yale, 1968) can be directly referred to:

> ### *2.1. Isometries and similarities*
>
> The structure of euclidean space is completely determined by the various properties of the distance between two points. Thus the two natural types of symmetries of euclidean space are those permutations that leave distances invariant, i.e., map a segment to a congruent segment, and those that preserve equal distances, i.e., map pairs of congruent segments to pairs of congruent segments. In this section we investigate the basic properties of the symmetries that preserve equal distances. In particular we show that a permutation of euclidean space preserves equal distances if and only if it stretches (or contracts) *all* distances by a constant factor, i.e., if and only if it corresponds to our intuitive concept of a similarity transformation.
>
> *Definition.* A permutation, ϕ, of euclidean space is called a *similarity* if for all points W, X, Y, and Z, $|W, X| = |Y, Z|$ if and only if $|(W)\phi, (X)\phi| = |(Y)\phi, (Z)\phi|$. In other words, a similarity is a permutation that *preserves equal distances.* If ϕ *preserves distances*, i.e., if $|W, X| = |(W)\phi, (X)\phi|$ for all W and X, then we call ϕ an *isometry* [from the Greek *isos* (equal) and *metron* (measure)].
>
> *Paul B. Yale, "Geometry and Symmetry" Holden-Day (1968) p46.*

As can be seen, the author uses the preservation of equal distances, thus skillfully covering scaling, translation, rotation and inversion operations. In my humble opinion, defining "similarity" via similarity transformations would also be a clear definition as it has its roots in crystallography. The sentence that starts with "If we have a metric" on line #53 could benefit from rewriting for clarification.

p3#41 Even though the reader is referred to the relevant article (Andrews et al. 2019a), The relation between the signs of the components and being Selling-reduced can be included here, along with a brief description of the procedure for reduction otherwise.

p3#54 Even though the reader is referred to the relevant article (Andrews et al. 2019b), a brief description of the reflections(/permutations) would be helpful.

p3#54 results *are* stored.

p3#55 The *storage* of … holds 24 entries *each* at that point.

p4#6 As the given counts are for maximum possible occasions (where there is no two overlapping transformed point), does the fact that two iterations (45876 entries) being sufficient imply that no new point have been obtained during the 2nd iteration? No emergence of any new point could have been asserted as a stopping criterion.

p4#29 As "modification" encompasses many operations, I suggest the title to be changed to "Why must the S6 vectors be scaled?". (Can I also recommend that, instead of scaling the vectors to the reference lattice's vector representation, all the vectors (including the reference) would be normalized thus would enable the comparison of lattices from different datasets directly? (For clarification: lattice vector used in this context: the position vector connecting the point in question to the origin)

p5#59 Spherical surface indicates a 3D space. Unless the C3 representation is being used (with complex and real parts separated, that is), it should be mentioned that we are still in 6D and the mentioned sphere is a 6D object.

p6#50 As the two structures have been found to be identical and this work focuses on lattices, not the atomic positions, either 1DPY or 1FE5 should be dropped as they will certainly yield the same results under transformations.

p6#55 In Tables 2, 3 *and 4* the first entry…

p8#15 Paragraph text has been placed in Table 5's caption

p8#59 The reference cell is given as C – shouldn't the unit cell parameters have changed if it was represented as a primitive lattice type?

p9#16 as only the endpoints are left, Table 9 is redundant (it's included in Table 10)

p12#6 It's not necessary to mention the omission of the space group information as they have not been included in anywhere else as well.

p12#46 Summary section should be enriched with a brief description of what has been achieved and outlook, if any. Being identical to its synopsis was a little bit limiting.

In general, it would be more helpful if the transformation matrices (to the listed transformed cells) have also been included in the comparison tables.

I also have a question (to be taken in the same manner as a post-presentation question – doesn't necessarily need to be addressed in the manuscript) to the authors: what is your opinion on the compliance of your method with the normalizers' restrictions upon the transformation matrices? As you are only considering the lattice parameters, I'm aware that you are not bounded by the restrictions of symmetries on atomic positions, but if one was to use your method for the comparison of structures, how would you advice on the implementation?

review from referee 2

Referee 2 attached the following notes to their review.

Disclaimer: I was asked to review this paper although I'm from a different discipline, but I was told it's okay if I just write about the parts that I can judge, and leave the rest to the other reviewers. The problem is that I'm lacking knowledge of the crystallography literature and can thus not really judge:

- (*a*) Is the work original and new?
- (*e*) Is the literature adequately and correctly referenced?
- (*f*) Is the journal for which the article has been submitted the most appropriate?

Summary of my opinion: From my perspective the paper seems good. The algorithm is explained well. I think the paper would be fine as it is, but maybe some details can even be improved further for the camera-ready version. I will list below things that I noticed. Feel free to ignore aspects where I am in the wrong.

Summary of the paper: The authors describe an algorithm how to change the unit cells of lattices (without changing the lattices) to become closest to a uniform scaling of a certain target unit cell. Apparently for this application the authors consider all uniform scalings (i.e. multiplying everything by a possibly large scale factor) of a lattice as equivalent to each other. This makes sense as their goal is measuring approximate similarity, not approximate equality of the lattices.
If I understand correctly, the authors do not consider the atoms in the motif of a crystal for this work, but instead only work with the lattice of the crystal, with the idea being to make the unit cells of the lattices look as nearly similar as possible, in order to then for example visually compare the motifs, already living in nearly similar unit cells.
In the end they give some examples with tables showing how well the unit cells fit to each other after the change.

Things I noticed:

Aspects the authors' research does not address (which seems okay in my opinion, because one doesn't need to solve all problems of the world at once). Maybe it's worth mentioning though that these questions are not addressed?
- distinguishing lattices that are multiples of each other by a large scale factor: It seems on purpose by the authors to approximate similarity, not approximate equality. But I (as a lay-person) wonder whether crystals/lattices should indeed be considered close to each other if they differ by a large scale factor.
- Identifying if the unit cell is doubled: If one crystal is described by a (say triclinic) unit cell and a motif inside, and the other crystal is described by a twice as large unit cell (just two copies of the first unit cell next to each other) with two copies of the motif inside, such that the two crystals are indeed the same, just differently represented, the proposed algorithm would not find out that those two crystals are the same. I've heard that this unintentional doubling of the unit cell can happen for example in crystal structure prediction, where the computer simulation might output a motif consisting of two almost perfect copies of a smaller motif, without realizing the pseudo-symmetry that would allow for the smaller unit cell. However, I can imagine that in many applications (maybe even the vast majority of applications) one can assume that this will not happen, so I understand why the authors do not cover this.
- no proof of exhaustiveness: There is no proof that the algorithm always gives a good fit between two nearly similar crystals. Instead the authors say that two iterations seemed to be sufficient in their experiments, but it's not clearly explained what "sufficient" means. Is there a ground truth one can compare it to? I guess it just means that in their experiments the fit did not considerably improve when the number of iterations was increased to more than 2. And I guess it's difficult to make that statement more precise. However, there is yet a different question, namely, is there a theoretical guarantee that the algorithm always would find the best possible fit after arbitrarily many iterations?

Aspects the paper maybe should address more:
- How to choose the reference cell? I feel it should be mentioned that Table 2,3,4 show that the fit is not symmetric, i.e. the fit between 1U4J and 1DPY differs (even by a factor 2 roughly) from the fit between 1DPY and 1U4J. This shows that it seems to matter how to choose the reference cell. However, I feel there is not much guidance given how to do that. Furthermore when choosing 1DPY as the reference cell, the fit values in Table 2, can't tell how close 1G0Z,1G2X,1U4J, and 2OSN are to each other. But I guess you anyway suggest, to use this method only as a first step, and then afterwards compare them visually, so I guess for that it doesn't matter too much which unit cell is the reference cell, right?
- What do the examples tell us? You describe the examples, but you don't describe what one can learn from them. I feel this is missing a bit.
- I found the last sentence of Section 6 pretty unclear, and would appreciate if it was better explained in the paper: "If it is desirable to restore lattice centering, then that operation must also be performed." Same with the last sentence of Section 7.3: "Because the initial cell was C-centered, the following cells are also in that presentation, although the intermediate cells are not C-centered." In general, I (as a lay-person) find it confusing what it even is supposed to mean that a Primitive unit cell is in C-centered presentation. I thought Primitive has less symmetry than C-centered. So how can something be in a more symmetric presentation than the symmetry that it has? But maybe that's only confusing to me.
- You give numbers for the "fit" but it's hard to judge whether these numbers should be considered as a rather good fit or a rather bad fit, given that one doesn't really have something to compare it to. For example, I'd be curious what's the worst possible fit (so largest number). But I guess it's not so easy to find out. But one could for example say that in Table 11, fit=35 appears, so I guess all the fits in Tables 2,3,4,6,7,8,13 which are all smaller than 6 should be considered rather small, so good fits?

Maybe in the abstract / introduction one could emphasize a bit more the following aspects:
- you allow for a scale factor
- you work with the lattice of the crystal, not the crystal itself, i.e. the motif inside the unit cell is irrelevant for this paper, just the unit cell parameters themselves matter.

Small comments:
- In the fourth line of Section 7 Examples, why do you say "two of which are identical", that sounds like there is one pair of two identical lattices. Don't you rather mean there are two pairs of each two identical lattices, namely lines 1, 2 and lines 3, 5 in Table 1? And wouldn't it make sense to remove line 2 and 5 if they are anyway identical copies of line 1 and 3?
- In Tables 5,6,7,8 it would help to have the rows numbered. Otherwise it's a bit hard to see which one of the two originally C-centered lattices is which.
- In Figure 1 and 2 you mention "regions closest to each of the points". You could mention that they are called Voronoi or Brillouin zones. But it's also perfectly okay not to mention this.
- Sometimes information could be given a bit more concise / dense, for example the beginning of the second paragraph of Section 7.3 is pretty similar to the paragraph beforehand.

Typos (I noticed them while reading, so why not let you know, but of course they are very minor)
- I think the "be" in "are be performed" in the 7th line of Section 3 should be removed or changed to "being" or "to be" or so, right?
- Is there a typo in the first line of the caption of Figure 2 "of the determining similarity" sounds a bit off to me, but maybe I'm wrong.
- In the sixth line of Section 7 Examples, you say "Tables 2 and 3" but I think you mean "Tables 2, 3, and 4", right?

review from referee 3

Journal: Acta Crystallographica Section A: Foundations and Advances
Paper: uv5018
Authors: Lawrence C. Andrews*, Herbert J. Bernstein and Nicholas K. Sauter
Title: Approximating Lattice Similarity

The manuscript by Andrews et al. introduces a potentially useful new algorithm for mapping a set of unit cell parameters such that they are as similar as possible to a reference unit cell. Their new algorithm is clearly presented, however section 7 ("Examples") would benefit from more detailed explanatory text. Section 7.2 is particularly sparse, with tables 5-8 not referenced in the main text at all. I would also welcome a (potentially brief) description of or comparison with existing alternatives to the described algorithm, and a comment on its numerical stability.

Page 1, line 57: "Data collections of unit cells" - the use of the term "data collections" is potentially confusing in this context as readers would likely more commonly associate the term with the act of collecting experimental data.
Page 2, line 46: This sentence might be clearer if it began "In Euclidean geometry, two objects ..."
Page 3, line 34: Delete "be" from "operations are be performed"
Page 4, line 8: "Two iterations" - should this be "Three iterations" based on the sequence of counts in the previous sentence?
Page 6, line 54: "In Tables 2 and 3" - I think this should be "In Tables 2, 3 and 4"
Page 7, Tables 2, 3 & 4: the authors aren't consistent in their choice of number of decimal places to report unit cell parameters
Page 8, section 7.2: I think this section needs more commentary on the tables. As is, tables 5-8 aren't referenced in the text at all.
Page 8, line 53: "Approximating a primitive cell" - should this be "Approximating a C-centered cell"?
Page 9, table 9: This table is not referenced in the text and could probably be removed from the paper - I don't believe it adds any information that isn't already in Table 10.
Page 9, line 16: Could the sentence beginning "The series of original points" be reworded to emphasise that this table is showing unit cells corresponding to a series of points in S6 rather than simply a "set of points" (and that the word "reduced" refers to "reduced unit cells").
Page 10, line 41: missing "." from final sentence.
Page 10, line 41: This sentence could be more explicit, e.g. "Table 13 lists the same cells in the same order after transformation to approximately match the reference cell."
Page 13, line 29: Extra whitespace in "(grant ..."