Metric spaces for comparing lattices; walking on a ripply surface in six or more dimensions

Lawrence C. Andrews[1] and Herbert J. Bernstein[2]

[1]Ronin Institute for Independent Scholarship, Kirkland, WA, USA, Lawrence.Andrews@RoninInstitute.org
[2]Ronin Institute for Independent Scholarship, c/o NSLS-2 BNL Bldg 745, PO Box 5000, Upton, NY 11973 USA, yayahjb@gmail.com

In 1978, we were asked to prepare an online database of all compounds with known unit cells so that users could search for unit cells of samples as a quick means of substance identification. At the time, there was no method to compute the "distance" between two unit cells (equivalently, lattices). There were two problems with the obvious metrics.

The first problem was that measured unit cell parameters (conventionally, $[a, b, c, \alpha, \beta, \gamma]$ for the cell lengths and angles) have experimental error in their determinations. Further, closely related compounds of interest might have slightly different parameters. That means that the problem to be solved is "the nearest neighbor problem" also know as "the post office problem". An exact match is not sufficient.

The second issue is related to the problem of experimental error, but it manifests in a different way. It is well-known that for any given lattice, there is an infinity of unit cells that can be chosen. Worse, in the presence of experimental error, no technique for selecting a "reduced cell" to uniquely represent each lattice can prevent two nearly identical lattices from having very different reduced cells. For example, in 1965 Lawton and Jacobson demonstrated two nearly identical lattices with very different Niggli reduced cells ($[6.490, 10.358, 10.359, 60.520, 71.750, 71.761]$ and $[6.490, 10.360, 10.369, 108.254, 108.246, 107.223]$) and in 1973 Gruber demonstrated a five-fold ambiguity in Buerger reduction.

This is not an abstract unphysical problem. Consider, for example, the unit cells of phospholipase A2 discussed by Le Trong and Stenkamp in 2007. They present three alternate cells from three different PDB entries that are actually for the same structure: $[57.98, 57.98, 57.98, 92.02, 92.02, 92.02]$ from entry 1FE5 by Singh *et al.* (2001) in space group R32, $[80.36, 80.36, 99.44, 90, 90, 120]$ from entry 1U4J by Singh *et al.* (2005) in space group R3 and $[80.949, 80.572, 57.098, 90.0, 90.35, 90.0]$ from entry 1G2X by Singh *et al.* (2004) in space group C2. No simple range search would recognize that these are the same.

If we have a well-defined metric for lattices, a helpful step is to search among "nearby" unit cells that differ from the probe cell. In 1978, the need for this approach was avoided by use of a perturbation stable 7-dimensional space, but, in general, searches with more sensitivity, especially near 90-degree angles, are needed.

Following our publication of the V7 algorithm in 1980 and our recognition of the deficiencies of the V7 algorithm in working near 90 degrees, we pursued an alternate solution using Niggli reduction (1988 and 2014). Among the problems that arise is that, in order to achieve a unique reduced cell, the unit cell edge lengths (a, b, c) are sorted by value and other inequalities are used. So if two parameters are equal, the question arises of how to sort given that experimental errors make all inequalities problematic. It is necessary to search among the "nearby" cells, and there may be quite a few. Our 2014 paper included the formal description of G^6, a metric space based on the parameters of Niggli reduction and also D^7, a new metric space.

The space G^6 is complex in the sense that the 15 boundaries of the fundamental unit are of several types. When we went on to consider Delaunay reduction, we immediately found that the space (we called it S^6) was vastly simpler than G^6. The fundamental unit in S^6 has only 6 boundaries, and they are all of the same type.

In addition, the light has finally dawned that it is counter-productive to try to force a unique reduced cell by heavy use of combinatorial inequalities. In the presence of experimental error, any choices made on such inequalities may flip. We are coming to the conclusion that achieving a good smooth metric requires us to accept the need for a reasonably high degree of ambiguity in the choice of a cell. In the case of S^6, we need 24 permutations and 6 boundary transforms.