A Useful Algorithm in Lattice Geometry

By D. J. Thomas

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England

(Received 12 September 1980; accepted 2 February 1981)

Abstract

An efficient algebraic algorithm is described for determining unit cells. It can be directly implemented on a digital computer. Two practical examples of its use are given: the determination of a unit cell on the (hkl) plane and the enumeration of the lattice points contained in a non-primitive cell. Both examples can be performed easily by hand; when performed by computer, they achieve a considerably higher computational speed than any other known methods.

It is often useful to find the largest primitive unit cell consistent with a given set of lattice vectors. An obvious case is the automatic determination of the unit cell of a reciprocal lattice from a limited number of diffraction spots. Two further examples which will be discussed below are the determination of the unit cell of the lattice which lies on the (hkl) plane and the enumeration of the lattice points which lie in a non-primitive cell. The former problem was discussed by Jaswon & Dove (1958) with reference to twin planes. The solution to the latter is useful in the control of area-detector diffractometers (Arndt & Gilmore, 1979).

The method described here for finding primitive unit cells is a generalization of Euclid's algorithm for finding highest common divisors (Hardy & Wright, 1979) extended to any number of dimensions. No such generalization appears to have been published before. To find the highest common divisor of a set of numbers by Euclid's algorithm, we repeatedly replace the largest number in the set by its remainder after division by the smallest non-zero one. When only one non-zero number remains, it is the highest common divisor. The remainder of one number after division by another is formally equivalent to the least non-negative residue of the first number modulo the second, expressed explicitly as

$$P = a \operatorname{frc}\left(\frac{p}{a}\right),$$

where P is the least residue of p modulo a and frc is a function whose value is the fractional part of its 0567-7395/81/040553-05\$01.00

argument, and is always taken to be positive: thus frc (3) = 0, frc $(4\frac{1}{4}) = \frac{1}{4}$, frc $(-5\frac{2}{3}) = \frac{1}{3}$.

This equation can be generalized to any number of dimensions in the obvious fashion:

$$\mathbf{P} = \mathbf{a} \operatorname{frc} \left(\mathbf{a}^{-1} \, \mathbf{p} \right), \tag{1}$$

where the columns of **a** represent the vectors which define the edges of a unit cell and **P** is the point inside that cell which is crystallographically equivalent to **p** and therefore related to it by a lattice vector (Fig. 1).

The matrix a^{-1} is the (generalized) inverse of **a**. It always exists and is unique, being defined (Penrose, 1955) by the equations:

$$\mathbf{a}\mathbf{a}^{-1}\,\mathbf{a} = \mathbf{a}, \quad \mathbf{a}^{-1}\,\mathbf{a}\mathbf{a}^{-1} = \mathbf{a}^{-1},$$

 $(\mathbf{a}\mathbf{a}^{-1})^{\dagger} = \mathbf{a}\mathbf{a}^{-1}, \quad (\mathbf{a}^{-1}\,\mathbf{a})^{\dagger} = \mathbf{a}^{-1}\,\mathbf{a},$

where $(\mathbf{a}\mathbf{a}^{-1})^{\dagger}$ represents the conjugate transpose of $\mathbf{a}\mathbf{a}^{-1}$. The normal matrix, $\mathbf{\tilde{a}}\mathbf{a}$, where $\mathbf{\tilde{a}}$ represents the transpose of \mathbf{a} will be non-singular for all matrices, \mathbf{a} , representing unit cells. Its general inverse is equal to the familiar inverse:

$$(\mathbf{\tilde{a}a})^{-1} = \frac{\operatorname{adj}(\mathbf{\tilde{a}a})}{\operatorname{det}(\mathbf{\tilde{a}a})}.$$

Therefore, the equation

$$a^{-1} = (aa)^{-1} a$$

is a suitable means for calculating a^{-1} in practice.



Fig. 1. \mathbf{P} and \mathbf{p} are crystallographically equivalent points since they are related by a lattice vector. The indices of \mathbf{P} are all between 0 and 1.

© 1981 International Union of Crystallography

With (1), Euclid's algorithm can be generalized to find a primitive unit cell of the lattice defined by a redundant set of vectors (Fig. 2).

The first step is to create the largest matrix, **a**, whose columns are linearly independent vectors chosen from the set. A primitive unit cell can then be found by the following steps.

[1] Replace one of the vectors which remains in the set but which does not form a column of **a** with its least residue (modulo **a**) as defined by (1). If that is the null vector, **0**, remove it from the set and go to step [2]. Otherwise exchange that residue with any column of **a** so that the columns remain linearly independent.

[2] If there are still any vectors in the set which do not form a column of **a**, go back to step [1]. The remaining vectors represent the edges of a primitive unit cell when they are all columns of **a**.

Two examples of the use of this algorithm are given. The first is an alternative approach to the twin-plane problem discussed by Jaswon & Dove (1958), which is more appropriate for automatic computation by a digital computer. It also has the advantage of working for all (hkl) planes, whereas Jaswon & Dove's method does not apply directly if h, k and l are not co-prime. The second example is a problem posed by the extremely high data-collection rates of a television type of area-detector diffractometer and the solution was especially designed for its high computational speed.

Example (1)

The determination of a unit cell of the lattice which lies on the (hkl) plane.

Lemma: A primitive unit cell of the lattice lying on the (hkl) plane can be found by applying the generalized form of Euclid's algorithm to the vectors:

$$\mathbf{h}^* = \frac{1}{(k|l)} \begin{bmatrix} 0\\ \bar{l}\\ k \end{bmatrix}, \quad \mathbf{k}^* = \frac{1}{(l|h)} \begin{bmatrix} l\\ 0\\ \bar{h} \end{bmatrix}, \quad \mathbf{l}^* = \frac{1}{(h|k)} \begin{bmatrix} \bar{k}\\ h\\ 0 \end{bmatrix}$$

where round parentheses containing numbers separated by vertical bars represent the greatest common divisor of those numbers.



Fig. 2. The redundant set of vectors **a**, **b**, **c** generate the lattice shown, but no two of them can be used as the edges of a primitive unit cell.

Proof: The (hkl) plane passing through the origin will always contain the vectors h^* , k^* , l^* , since

$$[h,k,l]$$
. $\mathbf{h}^* = 0$. $[h,k,l]$. $\mathbf{k}^* = 0$, $[h,k,l]$. $\mathbf{l}^* = 0$.

The area, S, of the unit cell of the lattice generated by, say, h^* and k^* alone is given by their vector cross product:

$$\mathbf{S} = \mathbf{h}^* \times \mathbf{k}^* = \frac{l}{(k|l)(l|h)} \begin{vmatrix} h \\ k \\ l \end{vmatrix},$$

but, as shown in the Appendix, the area of the primitive cell of the lattice on the (hkl) plane is given by

$$\mathbf{S}_p = \frac{1}{(h|k|l)} \begin{bmatrix} h \\ k \\ l \end{bmatrix},$$

which is smaller by the factor

$$\frac{|\mathbf{S}|}{|\mathbf{S}_n|} = \frac{l(h|k|l)}{(k|l)(l|h)} = r,$$

so that h^* and k^* do not generally form a primitive unit cell.

The equation

$$p\mathbf{h}^{*} + q\mathbf{k}^{*} + r\mathbf{l}^{*} \equiv \begin{bmatrix} 0 & l/(l|h) & \bar{k}/(h|k) \\ \bar{l}/(k|l) & 0 & h/(h|k) \\ k/(k|l) & \bar{h}/(l|h) & 0 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \end{bmatrix} = \mathbf{0}$$

describes a triangle of sides ph^* , qk^* and rl^* . The lowest whole-number solution is

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix} = \frac{1}{(h(k|l)|k(l|h)|l(h|k))} \begin{bmatrix} h(k|l) \\ k(l|h) \\ l(h|k) \end{bmatrix}$$

With the equation derived in the Appendix, this reduces to

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix} = \frac{(h|k|l)}{(h|k)(k|l)(l|h)} \begin{bmatrix} h(k|l) \\ k(l|h) \\ l(h|k) \end{bmatrix}$$

Hence, if, say, h^* and k^* are used to define a lattice, the smallest multiple of l^* which is a lattice point is

$$r\mathbf{l}^* = \frac{l(h|k|l)}{(k|l)(l|h)} \mathbf{l}^*.$$

The primitive cell of the lattice generated by \mathbf{h}^* , \mathbf{k}^* and \mathbf{l}^* is therefore smaller by the factor r than the unit cell of the lattice generated by \mathbf{h}^* and \mathbf{k}^* alone.

However, the ratio $|\mathbf{S}|/|\mathbf{S}_p|$ is the same multiple, r; thus the lattice which can be generated by \mathbf{h}^* , \mathbf{k}^* and \mathbf{l}^* contains every point on the (hkl) plane. The lemma stated above must follow since the application of Euclid's algorithm to these vectors must result in a primitive unit cell of the lattice they generate.

As a practical example, consider the (529) plane first discussed by Jaswon & Dove (Fig. 3). The vectors h^* , k^* , l^* are

$$\mathbf{h^*} = \begin{bmatrix} 0\\ \bar{9}\\ 2 \end{bmatrix}, \quad \mathbf{k^*} = \begin{bmatrix} 9\\ 0\\ \bar{5} \end{bmatrix}, \quad \mathbf{l^*} = \begin{bmatrix} \bar{2}\\ 5\\ 0 \end{bmatrix}.$$

To apply Euclid's algorithm, join **I*** and **h***, say, to form the matrix

$$\mathbf{a} = \begin{bmatrix} \bar{2} & 0\\ 5 & \bar{9}\\ 0 & 2 \end{bmatrix}.$$

Then the inverse of **a** is

$$\mathbf{a}^{-1} = \left\{ \begin{bmatrix} \bar{2} & 5 & 0 \\ 0 & \bar{9} & 2 \end{bmatrix} \begin{bmatrix} \bar{2} & 0 \\ 5 & \bar{9} \\ 0 & 2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} \bar{2} & 5 & 0 \\ 0 & \bar{9} & 2 \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{17}{44} & \frac{1}{22} & \frac{9}{44} \\ -\frac{9}{44} & -\frac{9}{110} & \frac{29}{220} \end{bmatrix}.$$



Fig. 3. A view of the (529) plane showing the vector **k**^{*}, which is equivalent to **k**^{*} in the lattice defined by **l**^{*} and **h**^{*}. It can be seen that this vector and either of **l**^{*} or **h**^{*} could be used as the edges of the primitive unit cell of the lattice on this plane.

For $\mathbf{p} = \mathbf{k}^*$, (1) becomes

$$\mathbf{k^{*\prime}} = \begin{bmatrix} \bar{2} & 0\\ 5 & \bar{9}\\ 0 & 2 \end{bmatrix} \operatorname{frc} \begin{bmatrix} -\frac{17}{44} & \frac{1}{22} & \frac{9}{44}\\ -\frac{9}{44} & -\frac{9}{110} & \frac{29}{220} \end{bmatrix} \begin{bmatrix} 9\\ 0\\ \bar{5} \end{bmatrix}$$
$$= \begin{bmatrix} \bar{2} & 0\\ 5 & \bar{9}\\ 0 & 2 \end{bmatrix} \operatorname{frc} \begin{bmatrix} -\frac{9}{2}\\ -\frac{5}{2} \end{bmatrix} = \begin{bmatrix} \bar{2} & 0\\ 5 & \bar{9}\\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2}\\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \bar{1}\\ \bar{2}\\ 1 \end{bmatrix}.$$

We now replace one of the columns of \mathbf{a} with $\mathbf{k}^{*'}$ and obtain

| | Ž | Ī | |
|------|---|---|--|
| a′ = | 5 | 2 | |
| | 0 | 1 | |

Repeating the application of (1) to the remaining vector, \mathbf{h}^* , gives $\mathbf{h}^{*\prime} = \mathbf{0}$; \mathbf{a}' therefore represents a primitive unit cell of the lattice lying on the (529) plane. This can be checked by observing that the cross product of the two columns of \mathbf{a}' is the same as the indices of that plane:

| $\bar{2}$ | | ī | | 5 | |
|-----------|---|---|---|---|--|
| 5 | x | 2 | = | 2 | |
| 0_ | | 1 | | 9 | |

Example (2)

The enumeration of the lattice points in a nonprimitive cell.

Lemma: The set of crystallographically distinct points of a lattice \mathscr{E} in a non-primitive cell α can always be found by enumerating the contents of another non-primitive cell, α^+ , whose index matrix, \mathbf{A}^+ , is diagonal (Fig. 4).



Fig. 4. A view down \mathbf{e}_3 of a supercell α and the orthogonal cell α^+ , containing equivalent lattice points. The numbered areas of the orthogonal cell are shown packed into the original supercell.

Proof: Consider a unit cell ε of a lattice \mathscr{E} whose defining edges are represented by the columns of a matrix **e** and also another unit cell α of a lattice \mathscr{A} whose volume is N times greater, represented by the matrix **a**. Suppose that every point of the lattice \mathscr{A} is a point of the lattice \mathscr{E} ; then α is a non-primitive cell of the lattice \mathscr{E} given by

$$\mathbf{a} = \mathbf{e}\mathbf{A},\tag{2}$$

where the index matrix **A**, whose determinant is N, is composed entirely of whole numbers. Given a vector whose indices are **p** in the \mathscr{A} lattice and **n** in the \mathscr{E} lattice, we can write

For all matrices that represent unit cells, this implies that

$$\mathbf{p} = \mathbf{A}^{-1} \mathbf{n}.$$

Now suppose that the first component of \mathbf{n} is N, the determinant of \mathbf{A} , and all the others are zero; then

$$\mathbf{n} = N\mathbf{1}_1,$$

where $\mathbf{1}_1$ is the first column of the unit matrix. Hence

$$\mathbf{p} = N\mathbf{A}^{-1} \mathbf{1}_1 = \operatorname{adj}(\mathbf{A}) \mathbf{1}_1,$$

which is the first column of the adjugate of A, since

$$\mathbf{A}^{-1} = \frac{\operatorname{adj} \mathbf{A}}{\operatorname{det} \mathbf{A}} = \frac{\operatorname{adj} \mathbf{A}}{N}$$

If the components of **p** have no common divisor higher than one, then there is no shorter vector in the \mathscr{A} lattice in the same direction. Thus all of the vectors $\lambda \mathbf{e}_1, \lambda = 1, 2, 3, ..., N$, where \mathbf{e}_1 is the first column of \mathbf{e} , will be crystallographically distinct in the \mathscr{A} lattice. Since there are N of them, they must represent all of the points of the lattice \mathscr{E} in the cell α .

If, however, **p** does have a common factor, say N', then the vector $(N/N') \mathbf{e}_1$ will be a vector of the \mathscr{A} lattice and only the vectors $\lambda \mathbf{e}_1$, $\lambda = 1, 2, 3, \dots, N/N'$, will be distinct, being repeated at intervals of $(N/N') \mathbf{e}_1$.

A view of the \mathscr{E} lattice in the direction of \mathbf{e}_1 is the same as a view of the lattice \mathscr{E}' of one less dimension, spanned by all the other vectors in \mathbf{e} . Similarly, a view of the \mathscr{A} lattice in the direction of \mathbf{e}_1 is the same as a view of the lattice \mathscr{A}' generated by the vectors in \mathbf{a}^* obtained by ignoring the \mathbf{e}_1 components of the vectors in \mathbf{a}^* lattice in one unit cell of the \mathscr{A}' lattice, we would then be able to find all of the points of the \mathscr{E}' lattice in one unit cell of the \mathscr{A}' lattice is one unit cell of the \mathscr{A}' lattice is none unit cell of the points of the \mathscr{E} lattice in one unit cell of the multiples of \mathbf{e}_1 to them. Thus, the problem has been reduced by one dimension and it is necessary to find the unit cell α' of the lattice \mathscr{A}' to make use of this fact. The projection of \mathscr{A} down \mathbf{e}_1 yields the same lattice, \mathscr{A}' , as can be generated by the projections \mathbf{a}^*

of the vectors **a** defining \mathscr{A} . Therefore, the vectors **a'**, defining the unit cell α' of the \mathscr{A}' lattice, can be found by applying Euclid's algorithm to the vectors in **a***.

To find the cell α' , create a new matrix e^* by replacing the first column of e with the null vector, 0, and also create a further new matrix, e', by simply removing the first column of e altogether. Then, e' is the unit cell of the lattice \mathcal{E}' of one less dimension than \mathcal{E} . The injections a^* of the vectors in a onto the sublattice \mathcal{E}' are given by an equation similar to (2):

$$a^* = e^* A.$$

Since a^* is an injection of a into one less dimension, it will have one column more than needed; Euclid's algorithm can be used to find the unit cell, a', of the lattice it generates. To find the points of the \mathcal{E}' lattice in a', we need the index matrix:

$$A' = e'^{-1} a'$$

We are now able to repeat the above procedure whence we shall obtain a further common factor, N'', and yet another index matrix of one less dimension still, namely:

$$\mathbf{A}^{\prime\prime}=\mathbf{e}^{\prime\prime-1}\,\mathbf{a}^{\prime\prime},$$

eventually stopping when the product of the common factors N', N'', \ldots is N, the total number of points of the \mathscr{E} lattice in the cell α . All subsequent common factors must be unity, so they do not need to be evaluated explicitly by the above procedure.

The contents of the orthogonal non-primitive cell a^+ , whose index matrix

$$\mathbf{A}^{+} = \begin{bmatrix} N/N' & 0 & 0\\ 0 & N'/N'' & 0\\ 0 & 0 & N''/N''' \end{bmatrix}$$

is therefore diagonal, can be enumerated directly by counting. They will not normally all lie in the same cell of the \mathscr{A} lattice, but can be grouped together by application of (1) (see Fig. 4).

As an example of this algorithm, consider the problem (shown in Fig. 4) of finding the lattice points which lie in the non-primitive cell whose edges are given by the columns of the index matrix

$$\mathbf{A} = \begin{bmatrix} 6 & 2 & 0 \\ \bar{2} & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Its adjoint is

$$adj(\mathbf{A}) = \begin{bmatrix} 4 & \bar{2} & 0 \\ 2 & 6 & 0 \\ 0 & 0 & 28 \end{bmatrix}$$

and its determinant is N = 28.

The first column, **p**, of the adjoint has a common divisor of N' = 2, so N/N' = 14. Taking the unit matrix as **e**, we get

$$\mathbf{e^*} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{e'} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which defines

$$\mathbf{a^*} = \mathbf{e^*} \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 2 & 0 \\ \bar{2} & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \bar{2} & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Applying Euclid's algorithm to a* gives:

$$\mathbf{a}' = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix},$$

which implies that

$$\mathbf{A}' = \mathbf{e}'^{-1} \mathbf{a}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

This has reduced the problem to two dimensions. Repeating it, we get

$$\operatorname{adj}(\mathbf{A}') = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$
 and $\operatorname{det}(\mathbf{A}') = N' = 2.$

The first column of the adjoint has no divisor greater than one, so N'' = 1, and N'/N'' = 2.

There is no need to go any further since N'' = 1, so N''' = 1. We now know all the terms necessary to define the cell a^+ represented by the index matrix A^+ :

$$\mathbf{A}^{+} = \begin{bmatrix} N/N' & 0 & 0 \\ 0 & N'/N'' & 0 \\ 0 & 0 & N''/N''' \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Examination of Fig. 4 will show that the points in a^+ do exactly fill the cell a when collected together.

The author gratefully acknowledges encouragement and advice from Drs U. W. Arndt, R. A. Crowther and A. Klug. The work was supported by a Postgraduate Training Award from the Medical Research Council of Great Britain.

APPENDIX

(1) The area of the unit cell on the (hkl) plane can be found simply by observing that the dot product of this area, \mathbf{S}_p , with the normal separation, \mathbf{R} , of the (hkl) planes is the volume of one primitive unit cell, that is unity. Thus, we have

$$\mathbf{S}_{p} \cdot \mathbf{R} = \mathbf{\tilde{S}}_{p} \mathbf{R} = 1,$$

which implies that

$$\mathbf{S}_p = \mathbf{\tilde{R}}^{-1}.$$

However, \mathbf{R}^{-1} is the shortest vector in the reciprocal lattice in the [h,k,l] direction. This is given by

$$\mathbf{R}^{-1} = \frac{[h,k,l]}{(h|k|l)},$$

r 7

which immediately gives the area required:

$$\mathbf{S}_p = \mathbf{\tilde{R}}^{-1} = \frac{1}{(h|k|l)} \begin{bmatrix} h \\ k \\ l \end{bmatrix}.$$

(2) The equation

$$(h|k|l)\left(h(k|l)|k(l|h)|l(h|k)\right) = (h|k)(k|l)(l|h)$$

can be proved as follows.

Any three numbers, h, k, l, can be factorized as $h = \alpha \beta_2 \beta_3 \gamma_1$, $k = \alpha \beta_3 \beta_1 \gamma_2$, $l = \alpha \beta_1 \beta_2 \gamma_3$, where $(h|k|l) = \alpha$, $(h|k) = \alpha \beta_3$, $(k|l) = \alpha \beta_1$, $(l|h) = \alpha \beta_2$. This implies that

$$(h|k|l)(h(k|l)|k(l|h)|l(h|k))$$

= $\alpha(\alpha\beta_2\beta_3\gamma_1\alpha\beta_1|\alpha\beta_3\beta_1\gamma_2\alpha\beta_2|\alpha\beta_1\beta_2\gamma_3\alpha\beta_3)$
= $\alpha^3\beta_1\beta_2\beta_3$
= $(h|k)(k|l)(l|h).$

References

- Arndt, U. W. & Gilmore, D. J. (1979). J. Appl. Cryst. 12, 1–9.
- HARDY, G. H. & WRIGHT, E. M. (1979). An Introduction to the Theory of Numbers, 5th ed., §§ 10.6 and 12.3. Oxford: Clarendon Press.
- JASWON, M. A. & DOVE, D. B. (1958). Acta Cryst. 8, 88-91.
- PENROSE, R. (1955). Proc. Cambridge Philos. Soc. 51, 406-413.