

Ambiguity of structure determination from a minimum of diffraction intensities

Ahmed Al-Asadi,^a Dimitri Leggas^{b,c} and Oleg V. Tsodikov^{b*}

^aUndergraduate Research Opportunities Program, University of Michigan, Ann Arbor, MI 48109, USA, ^bDepartment of Pharmaceutical Sciences, University of Kentucky, Lexington, KY 40536, USA, and ^cCarol Martin Gattton Academy of Mathematics and Science, Bowling Green, KY 42101, USA. Correspondence e-mail: ovt222@uky.edu

Although the ambiguity of the crystal structures determined directly from diffraction intensities has been historically recognized, it is not well understood in quantitative terms. Bernstein's theorem has recently been used to obtain the number of one-dimensional crystal structures of equal point atoms, given a minimum set of diffraction intensities. By a similar approach, the number of two- and three-dimensional crystal structures that can be determined from a minimum intensity data set is estimated herein. The ambiguity of structure determination from the algebraic minimum of data increases at least exponentially fast with the increasing structure size. Substituting lower-resolution intensities by higher-resolution ones in the minimum data set has little or no effect on this ambiguity if the number of such substitutions is relatively small.

© 2014 International Union of Crystallography

1. Introduction

A fundamental consequence of the loss of phase information in a diffraction experiment is that a crystal structure cannot be determined uniquely from intensities (I_{hkl}) alone, even if all I_{hkl} are known and error-free. This ambiguity was recognized many decades ago by Pauling (Pauling & Shappell, 1930) and Patterson (1944), but it has not been analyzed quantitatively. Around the same time, Ott (1927) and later Avrami (1938) formulated the problem of structure determination from a minimum of intensities in algebraic terms; however, how many structures can be obtained as a result is not known to this day, and no practical method of solving this algebraic problem has emerged yet. These original algebraic ideas were elegantly expanded more recently (Cervellino & Ciccariello, 1996), when the authors pinpointed the difficulties of developing a practical algebraic method due to the structure ambiguity, further exacerbated by experimental errors. Nevertheless, later they demonstrated that algebraic structure determination can be achieved for small structures of one-dimensional crystals, with added bond-length constraints and noted a large structure ambiguity (Cervellino & Ciccariello, 1999). Recently, by applying Bernstein's theorem to the problem of determining the structure of an idealized one-dimensional crystal from the algebraic minimum of intensities, we obtained the structure ambiguity for this oversimplified case of a one-dimensional crystal of identical point atoms as a function of the number of atoms in the unit cell (Al-Asadi *et al.*, 2012). Here, we apply Bernstein's theorem to establish this ambiguity for the realistic cases of three- and two-dimensional crystals.

2. Theoretical background

The determination of crystal structures of N equal point atoms in the unit cell with fractional coordinates (x_j, y_j, z_j) , $j = 1, 2, \dots, N - 1$ and $(x_N, y_N, z_N) = (0, 0, 0)$, after proper normalization, is equivalent to solving the following system of polynomial equations:

$$I_{hkl} = \left(1 + \sum_{j=1}^{N-1} \xi_j^h \eta_j^k \zeta_j^l \right) \left(1 + \sum_{j=1}^{N-1} \frac{1}{\xi_j^h \eta_j^k \zeta_j^l} \right), \quad (1)$$

where $\xi_j = \exp(2\pi i x_j)$, $\eta_j = \exp(2\pi i y_j)$, $\zeta_j = \exp(2\pi i z_j)$ and different equations (1) correspond to different choices of integer values of h , k and l . All $\xi_j = 1$ and all $\eta_j = \zeta_j = 1$ for two-dimensional and one-dimensional crystals, respectively. Therefore, for one-, two- and three-dimensional crystals, the minimum of $N - 1$, $2(N - 1)$ and $3(N - 1)$ error-free intensities, respectively, define a polynomial system that is, in principle, sufficient for structure determination. However, because this minimum system is composed of polynomial equations of high powers, it has many solutions, *i.e.* many structures yield the same intensities. Because of the sparsity of this system, the upper bound on the number of solutions given by the Bezout theorem (Newton *et al.*, 1729) vastly overestimates the number of its solution. A number of solutions of such sparse systems can instead be accurately estimated by the so-called mixed volume of the system, as stated in Bernstein's theorem (Bernstein, 1975). This approach is used here.

For a one-dimensional crystal of N atoms in a unit cell, the lowest resolution set of $N - 1$ intensities is constructed as I_h , $h = 1, 2, \dots, N - 1$, since the negative h correspond to the

Friedel pair counterparts. For two- and three-dimensional crystals, more than one index can increase in absolute value with increasing resolution. Whether an (h, k, l) triplet corresponds to an observable reflection is determined by the resolution, the geometry of the crystal lattice and the experimental setup. We will consider a generic scenario of a triclinic lattice with unit-cell parameters $a \simeq b \simeq c$ and $\alpha \simeq \beta \simeq \gamma \simeq 90^\circ$ and will assume that the data are complete to a given resolution limit. In this case, the resolution of a reflection (h, k, l) increases with $h^2 + k^2 + l^2$ (or to $h^2 + k^2$ for a two-dimensional crystal). We will also assume that the Friedel law holds and, to avoid including Friedel pairs into the data set, we will consider only reflections $\{h > 0, \text{any } k, l\}$, $\{h = 0, k > 0, \text{any } l\}$.

3. Results

3.1. The number of crystal structures obtained from the minimum of intensities when reflections lie on the axes of the reciprocal lattice

First, we will consider an idealized scenario when minimal system (1) is obtained for the reflections that lie on the reciprocal-lattice axes. Namely, the $3(N - 1)$ intensities of the minimum set are I_{h00} , I_{0k0} and I_{00l} , where $h, k, l = 1, 2, \dots, N - 1$. Equivalently, one obtains this case by omitting some low-resolution reflections I_{hkl} in the experimentally measurable set and instead including the above higher resolution ones on the reciprocal-lattice axes. As a consequence some equations in system (1) get replaced by equations of higher total degree (defined for each equation as $h + k + l$). This replacement leads to the increase in the number of possible solutions of system (1) (analogous to the increase of the number of solutions of a univariate polynomial equation as its degree increases), which is why the number of solutions in this idealized case is an upper bound. For this choice of reflections, the unknown coordinates (ξ_j, η_j, ζ_j) in system (1) separate:

$$I_{h00} = \left(1 + \sum_{j=1}^{N-1} \xi_j^h\right) \left[1 + \sum_{j=1}^{N-1} \left(1/\xi_j^h\right)\right], \quad h = 1, \dots, N - 1, \quad (2)$$

and the equations for I_{0k0} and I_{00l} have the same functional form as system (2). System (2) is equivalent to the one-dimensional crystal structure determination problem (Al-Asadi *et al.*, 2012; Shkel *et al.*, 2011). The number of all solutions of this system was recently obtained by our group (Al-Asadi *et al.*, 2012) as

$$n^{1D} = \frac{[2(N - 1)]!}{(N - 1)!}. \quad (3)$$

Because the solutions for each coordinate are independent of each other in this case, the numbers of solutions of equation (1) and its two-dimensional analog are equal to the cube and the square of the above number, respectively. To obtain the number of unique structures, one then needs to divide these numbers by $2N!$, to take into account the permutation of the atoms and the enantiomer obtained as a result of the center of

symmetry operation on each structure (for $N = 2$, these two operations are equivalent; therefore, in this case the factor of two is absent). We obtain

$$n_{\text{uniq,axes}}^{3D} = \frac{1}{2N!} \left\{ \frac{[2(N - 1)]!}{(N - 1)!} \right\}^3 \stackrel{N \gg 1}{\sim} \frac{(8N)^{2N}}{\exp(2N)}, \quad (4)$$

$$n_{\text{uniq,axes}}^{2D} = \frac{1}{2N!} \left\{ \frac{[2(N - 1)]!}{(N - 1)!} \right\}^2 \stackrel{N \gg 1}{\sim} \frac{(16N)^N}{\exp(N)}, \quad (5)$$

where the expressions for $N \gg 1$ are obtained by applying the Stirling approximation.

As explained at the beginning of this section, these factorial dependences of the numbers of unique structure solutions of system (2) generally greatly overestimate the numbers of unique structure solutions of system (1) obtained when the reflections do not necessarily lie on the reciprocal coordinate axes.

3.2. The lower bound on the number of crystal structures obtained from the minimum of intensities

To obtain a lower bound on the number of solutions of system (1), we will consider a practical case where reflections fill resolution shells completely with the increasing number of intensity data. We applied Bernstein's theorem (Bernstein, 1975) to system (1) by using an approach that was previously applied to a one-dimensional crystal (Al-Asadi *et al.*, 2012). The algebraic terminology and the formulation of Bernstein's theorem used in this study are the same as those published previously (Al-Asadi *et al.*, 2012) and, for convenience, are given again in the supporting information,¹ along with the technical details of the derivations. The so-called mixed volume, which we defined for this system previously (Al-Asadi *et al.*, 2012), is a good approximation of the total number of solutions of system (1), by Bernstein's theorem (Bernstein, 1975). The mixed volumes are, for a three-dimensional and a two-dimensional crystal, respectively (equations A11 and A12 in the supplementary material):

$$V_{\text{mix},3D} \geq \left\{ \frac{[2(N - 1)]!}{[(N - 1)!]^3} \right\}^3 (N - 1)! \frac{[3(N - 1)]!(9/2\pi)^{N-1}}{6^{N-1}}, \quad (6)$$

$$V_{\text{mix},2D} \geq \left\{ \frac{[2(N - 1)]!}{[(N - 1)!]^3} \right\}^2 (N - 1)! \frac{[2(N - 1)]!}{\pi^{N-1}}. \quad (7)$$

The number of unique structure solutions is obtained by dividing these values by $2N!$, where $N!$ is a number of permutations of coordinates of the N identical atoms in the structure and the factor of two accounts for the centrosymmetrical enantiomer ambiguity (except for $N = 2$, in which case the factor of two is absent, because these two operations are equivalent). We obtain:

¹ An appendix containing technical details of the derivations, the algebraic notation and the formulation of Bernstein's theorem discussed in this paper is available from the IUCr electronic archives (Reference: SC5073).

Table 1

Lower and upper bounds on the number of unique three- and two-dimensional crystal structures obtained from the minimum of lowest resolution diffraction data for $N \leq 20$.

N	Three-dimensional n_{uniq}^{3D} , lower bound†	$n_{\text{uniq,axes}}^{3D}$	Two-dimensional n_{uniq}^{2D} , lower bound‡	$n_{\text{uniq,axes}}^{2D}$
2	4§	4	2§	2
3	26§	144	6§	12
4	1576§	36000	56§	300
5	3860494	19756800	850§	11760
6	7.83×10^8	1.92×10^{10}	4357	635040
7	1.79×10^{11}	2.92×10^{13}	58612	43908480
8	4.46×10^{13}	6.42×10^{16}	836513	3.71×10^9
9	1.18×10^{16}	1.93×10^{20}	12481425	3.71×10^{11}
10	3.30×10^{18}	7.57×10^{23}	1.93×10^8	4.29×10^{13}
11	9.57×10^{20}	3.77×10^{27}	3.06×10^9	5.63×10^{15}
12	2.87×10^{23}	2.33×10^{31}	4.97×10^{10}	8.28×10^{17}
13	8.82×10^{25}	1.75×10^{35}	8.23×10^{11}	1.35×10^{20}
14	2.78×10^{28}	1.56×10^{39}	1.38×10^{13}	2.41×10^{22}
15	8.91×10^{30}	1.64×10^{43}	2.36×10^{14}	4.68×10^{24}
16	2.91×10^{33}	1.99×10^{47}	4.07×10^{15}	9.83×10^{26}
17	9.64×10^{35}	2.80×10^{51}	7.09×10^{16}	2.22×10^{29}
18	3.23×10^{38}	4.47×10^{55}	1.25×10^{18}	5.38×10^{31}
19	1.10×10^{41}	8.06×10^{59}	2.21×10^{19}	1.39×10^{34}
20	3.76×10^{43}	1.63×10^{64}	3.95×10^{20}	3.80×10^{36}

† The numbers are calculated from the expressions in inequality (10). ‡ The numbers are calculated from the expressions in inequality (11). § Because the accuracy of inequalities (10) and (11) decreases with decreasing N (as explained in the supporting information), the lower bounds for the smallest N were obtained from exact mixed volumes computed by *MixedVol* software (see Tables 2 and 3).

$$n_{\text{uniq}}^{3D} \geq \frac{1}{2N!} \left\{ \frac{[2(N-1)!]}{[(N-1)!]^3} \right\}^3 (N-1)! \frac{[3(N-1)]!(9/2\pi)^{N-1}}{6^{N-1}}, \tag{8}$$

$$n_{\text{uniq}}^{2D} \geq \frac{1}{2N!} \left\{ \frac{[2(N-1)!]}{[(N-1)!]^3} \right\}^2 (N-1)! \frac{[2(N-1)!]}{\pi^{N-1}}. \tag{9}$$

Finally, combining these lower bounds with the overestimate obtained in §3.1 for all consecutive reflections on the reciprocal axes into the same expression and applying the Stirling approximation at large N yields:

$$\frac{36^{2N}}{\pi^N} \stackrel{N \gg 1}{\sim} \frac{\{[2(N-1)!]^3 [3(N-1)!]\} \left(\frac{3}{4\pi}\right)^{N-1}}{2N! [(N-1)!]^8} \leq n_{\text{uniq}}^{3D} \leq \frac{1}{2N!} \left\{ \frac{[2(N-1)!]}{(N-1)!} \right\}^3 \stackrel{N \gg 1}{\sim} \frac{(8N)^{2N}}{\exp(2N)}, \tag{10}$$

$$\frac{8^{2N}}{\pi^N} \stackrel{N \gg 1}{\sim} \frac{\{[2(N-1)!]^2 [2(N-1)!]\} [2(N-1)!]}{2N! [(N-1)!]^5} \leq n_{\text{uniq}}^{2D} \leq \frac{1}{2N!} \left\{ \frac{[2(N-1)!]}{(N-1)!} \right\}^2 \stackrel{N \gg 1}{\sim} \frac{(16N)^N}{\exp(N)}. \tag{11}$$

Inequalities (10) and (11) show that the lower bound on the number of unique structures increases exponentially with N at large N , more strongly with increasing dimensionality (previously we obtained $n_{\text{uniq}}^{1D} \sim 4^N$ for large N ; Al-Asadi *et al.*, 2012). If reflections located only on the reciprocal axes are available, which is a simplifying, but not a realistic scenario,

then the number of unique solutions [the upper bound in inequalities (10) and (11)] increases faster than exponentially (factorially), underscoring the increase in ambiguity as one replaces lower-resolution reflections with higher-resolution ones. The numerical values for these bounds are given in Table 1 for $N \leq 20$. The ambiguity can be even larger, if low-resolution reflections are replaced by higher-resolution ones, as analyzed in §3.3.

3.3. Effect of replacing lower-resolution data by higher-resolution ones on ambiguity of crystal structure determination from a minimum intensity data set

Experimental data sets normally contain missing reflections; for example, the lowest-resolution reflections can be blocked by a beam-stop device. We investigated quantitatively how the ambiguity of structure determination changes as a result of increasing the resolution of the data set while keeping constant the number of data at its algebraic minimum. For different minimum data sets, we calculated the exact mixed volumes by using *MixedVol* software (Gao *et al.*, 2005; Li & Li, 2001) and obtained the number of unique three- and two-dimensional crystal structures (by dividing these values by $2N!$). These calculations were carried out for small N , as for larger N such calculations are computationally prohibitive. For each N , we first considered a set at a minimum resolution and then gradually increased the resolution of this set by substituting one or more of the lower-resolution reflections with higher resolution ones (Tables 2 and 3). These results demonstrate that a gain of a relatively few high-resolution reflections at the expense of a loss of low-resolution ones has only a minor effect on the ambiguity of crystal structure determination, and, in some cases, even reduces the ambiguity. An increase in ambiguity (if any) for such a small number of substitutions is much smaller than that for placing all the reflections on the reciprocal axes (Table 1, upper bound), or for increasing the structure size N .

4. Discussion

The problem of crystal structure determination from a minimum of experimental data was formulated in the algebraic form nearly a century ago (Ott, 1927), when neither analytical nor computational algebraic tools needed to deal with such treatment existed. With an advent of such tools in the last 15–20 years, we are in a position to analyze this problem and adapt these mathematical techniques to this specific problem. We showed previously that the ambiguity of solutions of polynomial system (1) for the one-dimensional case, when by definition all reflections lie on the reciprocal axis, increases exponentially. In this study, we analyzed more realistic cases of three- and two-dimensional crystals, where reflections can lie off the axes. Despite the fact that an efficient distribution of the reflections, when they gradually fill resolution shells from lowest to highest, is achievable in these cases, increase in ambiguity with increasing structure size is still, at best, exponential. Moreover, it becomes more dramatic

Table 2

Numbers of three-dimensional unique crystals structures calculated from the exact mixed volumes computed with *MixedVol* software for different sets of the algebraic minimum number of reflections.

Set	N	R^\dagger	Minimum reflection set (h, k, l)	V_{mix}	$n_{\text{uniq}}^{3D} = V_{\text{mix}}/(2N!)$
1	2	1	(100), (010), (001)	8	4‡
2		2	(110), (010), (001)	8	4‡
3		2	(100), (010), (011)	8	4‡
4		2	(100), (011), (001)	8	4‡
5	3	2	(100), (010), (001), (110), (101), (011)	312	26
6		2	(100), (010), (001), (110), (101), (011)	420	35
7		3	(100), (010), (001), (110), (101), (111)	312	26
8	4	3	(100), (010), (001), (110), (101), (011), (011), (111), (111)	93304	1943
9		4	(100), (010), (001), (110), (101), (011), (111), (200), (020)	79360	1653
10		5	(100), (010), (001), (110), (101), (011), (111), (200), (120)	82976	1728
11		6	(100), (010), (001), (110), (101), (011), (111), (211), (020)	75648	1576

† $R = h^2 + k^2 + l^2$, a parameter that increases with increasing resolution of the data. ‡ For $N = 2$, $n_{\text{uniq}}^{3D} = V_{\text{mix}}/N!$, as explained in the text.

Table 3

Numbers of two-dimensional unique crystals structures calculated from the exact mixed volumes computed with *MixedVol* software for different sets of the algebraic minimum number of reflections.

Set	N	R^\dagger	Minimum reflection set (h, k)	V_{mix}	$n_{\text{uniq}}^{2D} = V_{\text{mix}}/(2N!)$
1	2	1	(10), (01)	4	2‡
2		2	(10), (11)	4	2‡
3		2	(11), (11)	8	4‡
4		4	(10), (02)	8	4‡
5	3	2	(10), (01), (11), (11)	72	6
6		4	(10), (01), (11), (20)	72	6
7		5	(10), (01), (21), (20)	72	6
8		5	(10), (01), (11), (21)	72	6
9		5	(10), (21), (11), (20)	72	6
10	4	4	(10), (01), (11), (11), (20), (02)	3200	66
11		5	(10), (01), (11), (20), (02), (21)	3808	79
12		5	(10), (01), (12), (20), (02), (21)	7008	146
13		5	(10), (01), (11), (20), (12), (21)	3504	73
14		5	(10), (21), (01), (11), (11), (20)	2704	56
15		10	(10), (31), (11), (20), (02), (21)	5408	112
16	5	5	(10), (01), (11), (11), (20), (02), (21), (12)	204160	850
17		8	(10), (01), (11), (20), (02), (21), (12), (22)	227840	949
18		9	(10), (01), (11), (30), (02), (21), (12), (22)	341670	1423
19		10	(10), (31), (11), (20), (02), (21), (12), (22)	460160	1917
20		13	(10), (01), (11), (20), (02), (21), (32), (22)	279680	1165

† $R = h^2 + k^2$, a parameter that increases with increasing resolution of the data. ‡ For $N = 2$, $n_{\text{uniq}}^{2D} = V_{\text{mix}}/N!$, as explained in the text.

with increasing dimensionality. For a popular simplification, when the reflections lie on the reciprocal axes, increase in the ambiguity is even faster: it is factorial. These results demonstrate that, even if one can readily obtain all the structures that satisfy a minimum set of error-free intensities, finding the correct one from such a data set is not computationally feasible even for a reasonably small value of N . Does this mean that the algebraic approach should be dismissed? In attempting to answer this question one should take into consideration two key differences of this idealized problem from the practical experimental one, which offset each other: (1) the set of the intensities greatly overestimates the number of atomic coordinates (typically, 50–150-fold in small molecule

crystallography and 3–20-fold for macromolecular crystallography) and (2) intensity values contain experimental error and they do not satisfy exactly any one theoretical model, including system (1). Property (1) is key to reducing the ambiguity and, as demonstrated by traditional direct methods (Karle & Hauptman, 1957, 1958; Sheldrick, 2008; Weeks & Miller, 1999), allowing one to determine structures from intensities for a sufficiently small number of atoms and overcoming the increase in ambiguity due to property (2). The algebraic approach appears to have a potential of overcoming the restriction on the number of atoms, but it remains to be seen whether this potential is to be realised by modern computational methods. Ongoing research in our group is exploring these techniques towards developing new structure determination methodologies or improving existing ones.

References

Al-Asadi, A., Chudin, E. & Tsodikov, O. V. (2012). *Acta Cryst.* **A68**, 313–318.
 Avrami, M. (1938). *Phys. Rev.* **54**, 300–303.
 Bernstein, D. N. (1975). *Functional Anal. Appl.* **9**, 183–185.
 Cervellino, A. & Ciccariello, S. (1996). *Riv. Nuovo Cimento*, **19**, 1–59.
 Cervellino, A. & Ciccariello, S. (1999). *Z. Kristallogr.* **214**, 739–750.
 Gao, T., Li, T. Y. & Wu, M. (2005). *ACM T. Math. Software*, **31**, 555–560.
 Karle, J. & Hauptman, H. (1957). *Acta Cryst.* **10**, 515–524.
 Karle, J. & Hauptman, H. (1958). *Acta Cryst.* **11**, 264–269.
 Li, T. Y. & Li, X. (2001). *Found. Comput. Math.* **1**, 161–181.
 Newton, I., Motte, A. & Machin, J. (1729). *The Mathematical Principles of Natural Philosophy*. London: B. Motte.
 Ott, H. (1927). *Z. Kristallogr.* **66**, 136–153.
 Patterson, A. (1944). *Phys. Rev.* **65**, 195–201.
 Pauling, L. & Shappell, M. D. (1930). *Z. Kristallogr.* **75**, 128–142.
 Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
 Shkel, I. A., Lee, H. S. & Tsodikov, O. V. (2011). *Acta Cryst.* **A67**, 292–296.
 Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.