A Method for enumeration of powder auto-indexing solutions in Conograph

In the flowcharts of this section, every entry of an array $\Lambda^{obs} := \langle \langle q_i[0], q_i[1] \rangle : 1 \le i \le N_{peak} \rangle$ is a pair of a q-value $q_i[0]$ and its estimated error $q_i[1] = \mathrm{Err}[q_i[0]]$ of a peak. By using the powder auto-indexing method introduced here, the parameter of a lattice is provided as a metric tensor with entries $\frac{1}{4} \sum_{i=1}^{N_{peak}} c_i q_i[0]$ ($c_i \in \mathbb{Z}$), *i.e.,* linear sums of the q-values in Λ^{obs}.

At various stages of powder auto-indexing, the propagated errors of the linear sums of q-values are useful for making statistical judgments and strengthening the algorithm against observational errors. In order to simplify the notation, a formal sum $\sum_{i=1}^{N_{peak}} c_i q_i$ of the elements of Λ^{obs} is utilized in the flowcharts, and assumed to be equipped with the following functions getTerms, Val, Err and the less-than relation $<$:

$$\mathrm{getTerms}\left(\sum_{i=1}^{N_{peak}} c_i q_i\right) := \{q_i : 1 \le i \le N_{peak}, c_i \ne 0\}, \tag{A.1}$$

$$\mathrm{Val}\left(\sum_{i=1}^{N_{peak}} c_i q_i\right) := \sum_{i=1}^{N_{peak}} c_i q_i[0], \tag{A.2}$$

$$\mathrm{Err}\left(\sum_{i=1}^{N_{peak}} c_i q_i\right) := \left(\sum_{i=1}^{N_{peak}} c_i^2 (q_i[1])^2\right)^{1/2}, \tag{A.3}$$

$$\sum_{i=1}^{N_{peak}} a_i q_i < \sum_{i=1}^{N_{peak}} b_i q_i \underset{def}{\Leftrightarrow} \mathrm{Val}\left(\sum_{i=1}^{N_{peak}} a_i q_i\right) < \mathrm{Val}\left(\sum_{i=1}^{N_{peak}} b_i q_i\right). \tag{A.4}$$

If Val and Err are called with the argument $\sum_{i=1}^{N_{peak}} c_i q_i$, the value and propagated error of $\sum_{i=1}^{N_{peak}} c_i q_i[0]$ are returned, respectively.

A.1 Enumeration of powder indexing solutions

The algorithm of Table 1 enumerates powder indexing solutions in the array Ans. Although 3×3 metric tensors are constructed only from elements of Λ^{obs} satisfying the equation $3|l_1^*|^2 + |l_1^* + 2l_2^*|^2 = 3|l_2^*|^2 + |2l_1^* + l_2^*|^2$ in Oishi-Tomiyasu (2013a), Ito's equation is also utilized in Table 1, considering the cases in which the powder diffraction pattern contains only a small number of peaks.

Steps (1), (2) of Table 1 are based on the following assumptions respectively:

($H1$) If $q_r, q_s, q_t, q_u \in \Lambda^{obs}$ satisfy $2(q_r + q_s) = q_t + q_u$, then there exist $l_1^*, l_2^* \in L^*$ such that $q_r = |l_1^*|^2, q_s = |l_2^*|^2, q_t = |l_1^* + l_2^*|^2, q_u = |l_1^* - l_2^*|^2$.

($H2$) If $q_r, q_s, q_t, q_u \in \Lambda^{obs}$ satisfy $3q_r + q_t = 3q_s + q_u$, then there exist $l_1^*, l_2^* \in L^*$ such that $q_r = |l_1^*|^2, q_s = |l_2^*|^2, q_t = |l_1^* + 2l_2^*|^2, q_u = |2l_1^* + l_2^*|^2$.

In both quick and regular searches, the computation time of the procedure in Table 1 is roughly proportional to N_{zone}^2, where N_{zone} is the size of A_2 immediately after Step (2). This estimate is derived as follows: Steps (1)–(3) are much less time-consuming than Step (4), hence may be ignored. The size of A_3 in Step (3) is approximately $4N_{zone}$. When N_{peak} is the size of Λ^{obs}, the average number of $\langle R_1, R_2, R_3, R_4 \rangle \in A_3$ satisfying (a) or (b) with regard to fixed $\langle Q_1, Q_2, Q_3, Q_4 \rangle \in A_3$ is approximated with $4N_{zone}/N_{peak}$. Therefore, the number of combinations of $\langle Q_1, Q_2, Q_3, Q_4 \rangle$, $\langle R_1, R_2, R_3, R_4 \rangle \in A_3$ and $q_k \in \Lambda^{obs}$ is roughly equal to $4N_{zone} \cdot (4N_{zone}/N_{peak}) \cdot N_{peak} = 16N_{zone}^2$. Hence, the time is proportional to N_{zone}^2.

Table 1: Enumeration algorithm of powder indexing solutions.

void enumerate3DLattices($\Lambda^{obs}, c, \mathrm{MinDet}, \mathrm{MaxDet}, N_{sol}, Ans$)		
(Input)	Λ^{obs}	: array of N_{peak} pairs $\langle q_i[0], q_i[1]\rangle$ of a q-value $q_i[0]$ and its approximate error $q_i[1]$,
	$c > 0$	: parameter setting the error tolerance level,
	MinDet	: lower threshold for determinants of matrices in Ans,
	MaxDet	: upper threshold for determinants of matrices in Ans,
	N_{sol}	: upper threshold for number of entries in Ans.
(Output)	Ans	: array of 3×3 metric tensors.

(1) By the method in Table 2, enumerate q_r, q_s, q_t, q_u of Λ^{obs} satisfying $2(q_r + q_s) = q_t + q_u$ and insert $\langle \{q_r, q_s\}, \{q_t, q_u\}\rangle$ in A_2. (Here, A_2 is an array of four formal sums.)

(2) Enumerate q_r, q_s, q_t, q_u of Λ^{obs} satisfying $3q_r + q_t = 3q_s + q_u$. This is done by a method similar to that in Table 2. Using the new formal sum $q_{-1} := \frac{q_s+q_u-2q_r}{2} = \frac{q_r+q_t-2q_s}{2}$, two sets of q-values satisfying Ito's equation are generated:

$$2(q_{-1} + q_r) = q_s + q_u,\ 2(q_{-1} + q_s) = q_r + q_t. \tag{A.5}$$

Check whether A_2 contains $\langle \{q_w, q_r\}, \{q_s, q_u\}\rangle$ or $\langle \{q_w, q_s\}, \{q_r, q_t\}\rangle$ for some $1 \le w \le N_{peak}$. If not, this suggests that q_{-1} is the q-value of a diffraction peak undetected by peak search. Insert $\langle \{q_{-1}, q_r\}, \{q_s, q_u\}\rangle$ and $\langle \{q_{-1}, q_s\}, \{q_r, q_t\}\rangle$ into A_2.

(3) For each entry $\langle \{Q_1, Q_2\}, \{Q_3, Q_4\}\rangle$ of A_2, insert the following into a new array A_3:

$$\langle Q_1, Q_2, Q_3, Q_4\rangle, \langle Q_1, Q_2, Q_4, Q_3\rangle, \langle Q_2, Q_1, Q_3, Q_4\rangle, \langle Q_2, Q_1, Q_4, Q_3\rangle. \tag{A.6}$$

(4) For each $\langle Q_1, Q_2, Q_3, Q_4\rangle$ in A_3, search for $\langle R_1, R_2, R_3, R_4\rangle \in A_3$ satisfying either of the following:

(a) $Q_1 = R_1 \in \Lambda^{obs}$.

(b) $Q_1, R_1 \notin \Lambda^{obs}$ and $|\mathrm{Val}(Q_1 - R_1)| \le c\mathrm{Err}(Q_1 - R_1)$.

In addition, for every $q_k \in \Lambda^{obs}$, assume that there exist l_1^*, l_2^*, l_3^* satisfying

$$\begin{aligned} Q_1 \approx R_1 = |l_1^*|^2,\ Q_2 = |l_2^*|^2,\ Q_3 = |l_1^* + l_2^*|^2, \\ R_2 = |l_3^*|^2,\ R_3 = |l_1^* + l_3^*|^2,\ q_k = |l_1^* + l_2^* + l_3^*|^2. \end{aligned} \tag{A.7}$$

Then, the metric tensor $S := (l_i^* \cdot l_j^*)_{1 \le i,j \le 3}$ is obtained as the following 3×3 symmetric matrix:

$$\begin{pmatrix} Q_1 & \frac{Q_3-Q_1-Q_2}{2} & \frac{R_3-Q_1-R_2}{2} \\ \frac{Q_3-Q_1-Q_2}{2} & Q_2 & \frac{Q_1-Q_3-R_3+q_k}{2} \\ \frac{R_3-Q_1-R_3}{2} & \frac{Q_1-Q_3-R_2+q_k}{2} & R_2 \end{pmatrix}. \tag{A.8}$$

The values and propagated errors of the entries are computed using the functions Val and Err. If $\mathrm{MinDet} \le \det S \le \mathrm{MaxDet}$ holds, insert S into Ans in ascending order of $\det S$.[a] If the size of Ans exceeds N_{sol}, remove the last entry of Ans.

[a] In a regular search, this is replaced by insertion of S into Ans in descending order of the figure of merit proposed in Wu (1988).

Table 2: Enumeration algorithm of four q-values satisfying Ito's equation.

void enumerateItoEquations(Λ^{obs}, c, Ans)

(Input)	Λ^{obs}, c	: the same as in Table 1.
(Output)	Ans	: array of a sequence $\langle \{q_r, q_s\}, \{q_t, q_u\} \rangle$, where q_r, q_s, q_t, q_u are elements of Λ^{obs} satisfying

$$\begin{cases} |\mathrm{Val}(2q_r + 2q_s - q_t - q_u)| \le c \min\{2\mathrm{Err}(q_r + q_s), \mathrm{Err}(q_t + q_u)\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{(Ito's equation)} \\ \left(\frac{q_t[0] - q_r[0] - q_s[0]}{2}\right)^2, \left(\frac{q_u[0] - q_r[0] - q_s[0]}{2}\right)^2 \le q_r[0]q_s[0]. \\ \qquad\qquad\qquad\qquad\qquad \text{(positive definiteness)} \end{cases}$$

Set a sorted sequence $S := \langle q_i + q_j : 1 \le i \le j \le N_{peak} \rangle$ of formal sums.
for $i := 1$ to $\frac{1}{2} N_{peak}(N_{peak} + 1)$ do
  Let $1 \le J_{min}, J_{max} \le N_{peak}$ be integers satisfying
  $J_{min} \le j \le J_{max} \iff |\mathrm{Val}(2S[i] - S[j])| \le 2c\mathrm{Err}(S[i])$.
  for $j := J_{min}$ to J_{max} do
    if $|\mathrm{Val}(2S[i] - S[j])| \le c\mathrm{Err}(S[j])$ then
      $\{q_r, q_s\} := \mathrm{getTerms}(S[i])$,
      $\{q_t, q_u\} := \mathrm{getTerms}(S[j])$.
      Insert $\langle \{q_r, q_s\}, \{q_t, q_u\} \rangle$ in Ans,
      if q_r, q_s, q_t, q_u satisfy the condition for positive definiteness.
    end if
  end for
end for

A.2 Speed-up of the enumeration method using topographs

As described in Section A.1, the computation time of Table 1 is proportional to the square of the size N_{zone} of A_2 immediately after Step (2). Thus, an effective way to speed up the algorithm is to reduce the size of A_2.

Every entry $\langle \{R_1, R_2\}, \{R_3, R_4\} \rangle$ of A_2 consists of four formal sums satisfying Ito's equation $2(R_1 + R_2) = R_3 + R_4$, therefore corresponds to an edge of a topograph as in Figure 1. (See Oishi-Tomiyasu (2013a) for more detailed explanations about topographs.)

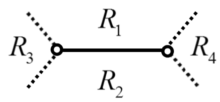

Figure 1: The edge of a topograph corresponding to Ito's equation $2(R_1+R_2) = R_3+R_4$.

In Oishi-Tomiyasu (2013a), a figure of merit for the entries of A_2 (that is, zones) was defined as the size of the graph T formed from entries of A_2. Table 3 presents the detailed algorithm to expand T starting from $\langle \{R_1, R_2\}, \{R_3, R_4\} \rangle$.

In Table 3, the graph in Figure 1 is expanded to only one side. In order to obtain the whole T and compute the figure of merit of $e := \langle \{R_1, R_2\}, \{R_3, R_4\} \rangle \in A_2$, it is necessary to call the recursive procedure twice with $\tilde{e} := \langle \{R_1, R_2\}, R_3, R_4 \rangle$ and $\langle \{R_1, R_2\}, R_4, R_3 \rangle$, and connect the two output graph as in Figure 3. The figure of merit $C(e)$ is then computed as the number of q-values in Λ^{obs} that appear in T. As explained in Oishi-Tomiyasu (2013a), e with larger $C(e)$ should be given priority among

Table 3: Recursive procedure to form a graph from $\langle\{R_1, R_2\}, \{R_3, R_4\}\rangle$ and entries of A_2.

void expandSubtopograph(A_2, $\tilde{e}$, $\tilde{A}_2$, T **)**

(Input) Λ^{obs} : the same as in Table 1,

A_2 : array of four formal sums $\langle\{Q_1, Q_2\}, \{Q_3, Q_4\}\rangle$ assumed that every entry satisfies $\mathrm{Val}(2Q_1 + 2Q_2 - Q_3 - Q_4) \le c \min\{2\mathrm{Err}(Q_1 + Q_2), \mathrm{Err}(Q_3 + Q_4)\}$ for some fixed constant c. Furthermore, it is assumed that Q_3, Q_4, and at least one of Q_1, Q_2 belong to Λ^{obs} (*i.e.,* there is a $q \in \Lambda^{obs}$ such that $Q_i = q$).

$\tilde{e}$: supposed to equal $\langle\{R_1, R_2\}, R_3, R_4\rangle$.

(Output) $\tilde{A}_2$: array containing entries of A_2.

T : graph[a] composed of edges corresponding to $\langle\{Q_1, Q_2\}, \{Q_3, Q_4\}\rangle$ in $\tilde{A}_2$.

For $(i, j) = (1, 2), (2, 1)$, let S_{ij} be the set defined by

$$S_{ij} := \begin{cases} \{\langle\{R_i, R_4\}, R_j, q\rangle \in A_2 : q \in \Lambda^{obs}\} & \text{if } R_j \in \Lambda^{obs}, \\ \emptyset & \text{otherwise.} \end{cases}$$

$T_{12} := \emptyset$. $A_{12} := \emptyset$. $T_{21} := \emptyset$. $A_{21} := \emptyset$.
for $i = 1$ to 2 do
    Take $j = 1, 2$ such that $j \neq i$.
    for $\tilde{e}_2 \in S_{ij}$ do
        Call expandSubtopograph(A_2, $\tilde{e}_2$, $\tilde{A}_{ij}$, $\tilde{T}_{ij}$).
        if $|A_{ij}| < \left|\tilde{A}_{ij}\right|^b$ then
            $T_{ij} := \tilde{T}_{ij}$.
            $A_{ij} := \tilde{A}_{ij}$.
        end if
    end for
end for
Set $\tilde{A}_2 := \{\langle\{R_1, R_2\}, \{R_3, R_4\}\rangle\} \cup A_{12} \cup A_{21}$.
Construct T by unifying T_{12}, T_{21} and the subgraph corresponding to $\langle\{R_1, R_2\}, \{R_3, R_4\}\rangle$, as shown in Figure 2.

[a] In the actual algorithm, the construction and the output of the graph T can be omitted, because the figure of merit for entries of A_2 is defined only from $\tilde{A}_2$.

[b] For any subset $\tilde{A}_2$ of A_2, the number of q-values in Λ^{obs} that appear in some entry of $\tilde{A}_2$ is denoted by $\left|\tilde{A}_2\right|$. (The $\tilde{A}_{ij}$ and $\tilde{T}_{ij}$ with larger $\left|\tilde{A}_{ij}\right|$ is given priority.)

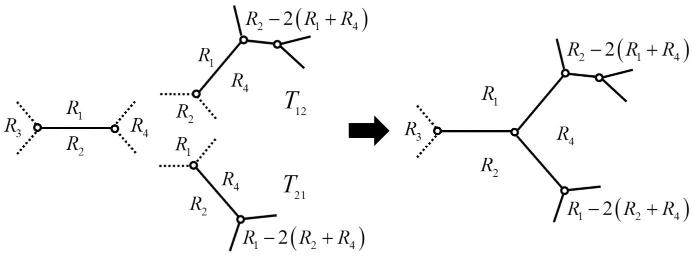

Figure 2: Extension of a topograph (1/2).

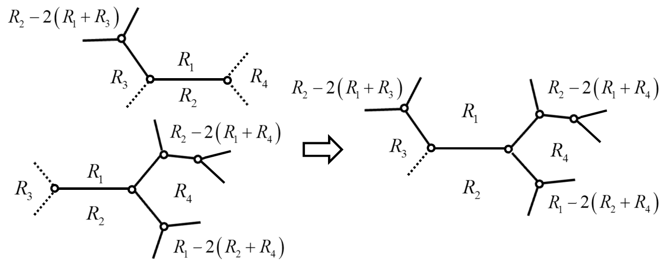

Figure 3: Extension of a topograph (2/2).

the entries of A_2.

In order to use $C(e)$ as a sorting criterion for zones, the following are inserted between Steps (2) and (3) of Table 1:

(2i) For each $e := \langle \{R_1, R_2\}, \{R_3, R_4\} \rangle$ in A_2, compute $C(e)$ by calling the procedure in Table 3.

(2ii) Sort A_2 in descending order of $C(e)$. For $e_1, e_2 \in A_2$ satisfying $C(e_1) = C(e_2)$, the sorting is done according to $e_1 < e_2 \underset{def}{\Longleftrightarrow} \det S(e_1) < \det S(e_2)$, where $S(e)$ is the following 2×2 metric tensor defined for $e := \langle \{R_1, R_2\}, \{R_3, R_4\} \rangle \in A_2$:

$$S(e) := \begin{pmatrix} \mathrm{Val}(R_1) & \frac{1}{2}\mathrm{Val}(R_3 - R_1 - R_2) \\ \frac{1}{2}\mathrm{Val}(R_3 - R_1 - R_2) & \mathrm{Val}(R_2) \end{pmatrix}.$$

(2iii) Remove the $(N_{zone}+1)$-th-to-last elements of A_2 using a fixed integer $N_{zone} > 0$. (The default value of N_{zone} used in Conograph is given in (A.10) of Section B.)

These procedures finish in a moment when the number of q-values in Λ^{obs} is below 100. In (2ii), $e \in A_2$ with smaller $\det S(e)$ is given priority, because Step (4) of Table 1 assumes that $\{l_1^*, l_i^*\}$ $(i = 2, 3)$ is a primitive set (*i.e.,* a subset of some basis) of L^*, hence their corresponding entries $\langle \{Q_1, Q_2\}, \{Q_3, Q_4\} \rangle, \langle \{R_1, R_2\}, \{R_3, R_4\} \rangle$ are considered to have a comparatively small determinant. It should be noted that $C(e)$ also has the property of ranking $e \in A_2$ with smaller $\det S(e)$ higher, because $\Lambda^{obs} \subset [q_{min}, q_{max}]$ contains a larger number of ${}^t u S(e) u$ $(u \in \mathbb{Z}^2)$ if $S(e)$ has a smaller determinant.

B Default setting of input parameters in Conograph

The character string AUTO is used in the Conograph software to generate default parameters depending on respective powder diffraction patterns. In what follows, we explains the formulas compute the value of AUTO.

1. **Number of q-values used.** After sorting the q-values in Λ^{obs} into ascending order, the $(N_{peak} + 1)$-th-to-last parameters are removed before powder auto-indexing commences. The default value of N_{peak} is calculated as

$$N_{peak} \quad := \quad \min \left\{ \sharp\{ q < 10/d^2 : q \in \Lambda^{obs} \}, 48, M \right\}, \qquad \text{(A.9)}$$

where $\sharp T$ is the number of elements of the set T, d is the lower threshold for the distance between two lattice points, and M is the number of peaks detected by peak search. (In the initial configuration of Conograph, d is set to $2\mathring{A}$.)

2. **Upper threshold for the number of $\langle\{Q_1, Q_2\}, \{Q_3, Q_4\}\rangle$.**

$$N_{zone} := \begin{cases} \frac{1}{3} N_{peak}(N_{peak}+1) & \text{in quick search,} \\ \frac{1}{2} N_{peak}(N_{peak}+1) & \text{in regular search.} \end{cases} \tag{A.10}$$

3. **Upper threshold for the number of candidate solutions.**

$$N_{sol} := \begin{cases} \min\{64000, N_{zone}^2\} & \text{in quick search,} \\ \min\{32000, 2N_{peak}(N_{peak}+1)(N_{peak}+2)/3\} & \text{in regular search.} \end{cases} \tag{A.11}$$

4. **Thresholds for the volume of the primitive cell.**

$$\mathrm{Vol}_{min} \quad := \quad \max\{5, v_{20}^{-1}\}, \tag{A.12}$$

$$\mathrm{Vol}_{max} \quad := \quad 30\mathrm{Vol}_{min}, \tag{A.13}$$

where $5\,\mathring{A}^3$ is chosen as the lower threshold for the volume of existing crystals. The parameter v_{20} is an upper bound on the volume of the reciprocal cell $\mathbb{R}^3/L^*$, which is estimated using the 20 smallest q-values of $\Lambda^{obs} := \{q_1, q_2, \ldots, q_{N_{Peak}}\}$ as follows:

$$v_j := \frac{2\pi}{3} \frac{q_j^{3/2} - q_1^{3/2}}{j-1}. \tag{A.14}$$

The numbers 20 and 30 in (A.12), (A.13) were adopted empirically. If $N_{Peak} < 20$, $v_{N_{Peak}}$ is used instead of v_{20}.