Short note: analysis of the non-uniqueness of seismic travel-times through brute-force counting

Kris Innanen

ABSTRACT

The questions of how "hard" seismic inversion is, and of how good a solution we can expect, both rest on the uniqueness of a seismic datum (such as a travel time), and the degree to which multiple independent data of the same type can mitigate it. Since large numbers of unknowns are involved, one might think to use the methods of statistical mechanics to characterize this non-uniqueness. Anyway, I did. Statistical analyses usually start with a counting of possibilities (e.g., the number of states of a many-particle system with the same energy). In our case the counting would be of the number of the possible models which produce the same traveltime (or amplitude, or waveform component, etc.). How precisely can we count the number of discrete slowness models which produce a given traveltime? Here we will set out an approach in a short note. Once we have a method for counting, we may be able to glean interesting facts about the inverse problem in a range of experiments and seismic source/receiver configurations.

COUNTING TRAVELTIME-PRESERVING SLOWNESS MODELS

How many velocity models are there which produce the same seismic traveltime along a fixed path? Let us broach this in steps. First, we will re-write the traveltime-slowness relationship in a way which enables counting. Then we will relate the traveltime calculation to the problem of determining partitions. Finally, we will show that if we can count partitions, and if we can also count the unique re-orderings of partitions, we can count all of the velocity models that can lead to a given traveltime.

1. A combined discretization of slowness and path

The traveltime of a wave propagating at speed c a distance L along a path with differential length elements dl is

$$\tau = \int_0^l \frac{dl'}{c(l')} = \int_0^l s(l')dl', \tag{1}$$

where s is the slowness. Expanding s over a pulse basis allows us to treat the problem in terms of a sequence of N real and positive numbers s_i:

$$s(l) = \sum_{i=1}^{N} s_i \Big[H(l - l_i) - H(l - l_{i-1}) \Big] \tag{2}$$

where H is the Heaviside function

$$H(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases} . \tag{3}$$

Upon substitution of equation (2) into (1) we obtain

$$\tau = \Delta l \sum_{i=1}^{N} s_i, \tag{4}$$

where $\Delta l = l_i - l_{i-1}$ is the (assumed constant) width of the pulse. This discretizes the path, with each element of the sequence $\{s_1, s_2, ..., s_{N-1}, s_N\}$ being associated with a distance along the path of the wave $\{l_1, l_2, ..., l_{N-1}, l_N\}$. The slownesses themselves have not been discretized, however, since each s_i can take on any value. Let us discretize them also. First, we define

$$S = \frac{\tau}{\Delta l}, \tag{5}$$

as the largest slowness* we can assign to any s_i if we are given a τ and an N. Then, we ask that each s_i be selected from a set of $N + 1$ discrete, regularly spaced values between 0 and S:

$$\{0, \Delta s, 2\Delta s, ..., S - 2\Delta s, S - \Delta s, S\}, \tag{6}$$

with $\Delta s = S/N$. Lastly, let us change the system of units we will use in all of our forthcoming calculations:

$$s_{\text{new units}} = s_{\text{old units}} \times \left(\frac{N}{S}\right). \tag{7}$$

In this system, by equations (4), (5), and (6), any model s_i^* which preserves τ must satisfy

$$\sum_{i=0}^{N} s_i^* = N, \quad s_i^* \in \{0, 1, 2, ..., N\}. \tag{8}$$

This is a general expression, since N can be chosen large enough to attain any desired accuracy in either the slowness or path approximations. Thus it is possible to write the traveltime-slowness relationship for any fixed path, with any desired degree of accuracy, in the form in equation (8). See Figure 1. The number N is not fixed - eventually in a statistical analysis however we will make progress by assuming N is large (such that, e.g., $N!$ can be replaced with Stirling's approximation).

2. Partitions

The importance of equation (8) is that it establishes a one-to-one connection between a discretized slowness-traveltime relation and the partitions of an integer N. For example, take $N = 5$. To satisfy the relation, s_i^* must contain a set of 5 non-negative integers, up to and including 5, whose sum is 5. There are seven unique (i.e., not re-ordered) sequences of

*If any one s_i value in equation (4) is S, all others must be zero to preserve τ.

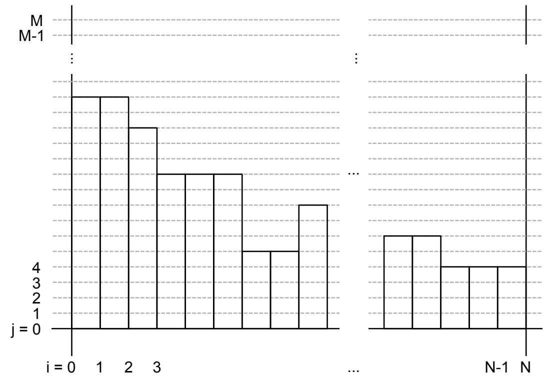

FIG. 1. Discretized slowness model. Any velocity model between the depths 0 and Z can be approximated by N layers each of which takes on one of M slowness values. In our analysis, in order to make use of partitions, we will need to enforce $M = N$. Provided N and M are large, this does not meaningfully affect the generality of the results. In any particular case $N = M$ can be chosen large enough to be adequate for whichever of the space or slowness discretization has the more stringent requirements.

this kind:

$$\begin{aligned} &\{5, 0, 0, 0, 0, 0\} \\ &\{4, 1, 0, 0, 0, 0\} \\ &\{3, 2, 0, 0, 0, 0\} \\ &\{3, 1, 1, 0, 0, 0\} \\ &\{2, 2, 1, 0, 0, 0\} \\ &\{2, 1, 1, 1, 0, 0\} \\ &\{1, 1, 1, 1, 1, 0\}. \end{aligned} \tag{9}$$

These are well-known in number theory as the 7 *partitions* of 5. In fact, equation (8) is simply an equation constraining allowable sequences of slowness values s_i^* to those which can be written as partitions of N. Counting traveltime preserving velocity models is therefore related to counting the partitions of N, i.e., $p(N)$ where for instance $p(5) = 7$.

3. Partition re-ordering

The partitions in equation (9) do not represent a full enumeration of all sequences whose sum is 5. We can re-order the top row, for instance, exchanging the 5 with any 0 and get the same sum. To get an exhaustive list of all the sequences summing to 5, we also need to count up all of the ways each row of equation (9) can be re-ordered. In general there are $Q!$ re-orderings of Q elements. A set with $Q = 1$ element, say $\{a\}$, has 1 ordering. A set with $Q = 2$ elements, say $\{a, b\}$, has 2 different orderings:

$$\{a, b\}, \quad \{b, a\}. \tag{10}$$

When we move to $Q = 3$ elements (and higher), the number of re-orderings grows and the trick becomes to find a regular way of producing them so they can be counted. Here is one way. First, put a ‘c’ in the front, middle, or back of the first $Q = 2$ ordering, $\{a, b\}$:

$$\{c, a, b\}, \quad \{a, c, b\}, \quad \{a, b, c\}. \tag{11}$$

Then, put the 'c' in the front, middle, or back of the second ordering, $\{b, a\}$:

$$\{c, b, a\}, \quad \{b, c, a\}, \quad \{b, a, c\}. \tag{12}$$

That is all of the possibilities. So, each of the two $Q = 2$ cases can produce 3 new orderings, for a total of 6. For $Q = 4$, we repeat this process; each of the 6 sequences from the $Q = 3$ case can accommodate a 'd' in 4 spots, the front, front middle, back middle, and back locations, for a total of 24 new orderings:

$$\begin{array}{llll}
\{d,c,a,b\}, & \{c,d,a,b\}, & \{c,a,d,b\}, & \{c,a,b,d\} \\
\{d,a,c,b\}, & \{a,d,c,b\}, & \{a,c,d,b\}, & \{a,c,b,d\} \\
\{d,a,b,c\}, & \{a,d,b,c\}, & \{a,b,d,c\}, & \{a,b,c,d\} \\
\{d,c,b,a\}, & \{c,d,b,a\}, & \{c,b,d,a\}, & \{c,b,a,d\} \\
\{d,b,c,a\}, & \{b,d,c,a\}, & \{b,c,d,a\}, & \{b,c,a,d\} \\
\{d,b,a,c\}, & \{b,d,a,c\}, & \{b,a,d,c\}, & \{b,a,c,d\}
\end{array} \tag{13}$$

For $Q = 3$ we therefore have $3 \times 2 = 6$ orderings, and for $Q = 4$ we have 4 times this many, or $4 \times (3 \times 2) = 24$; the factorial rule in this way becomes evident. With this result, the problem of counting the number of total velocity models which preserve the traveltime is evidently easy. Each partition of N has $N!$ re-orderings, and there are $p(N)$ partitions, so the total number of models with the same τ is

$$n_\tau = N!\, p(N). \tag{14}$$

This can be re-written in a less compact way as

$$n_\tau = \sum_{i=1}^{p(N)} N!, \tag{15}$$

in which we (obtusely) add an $N!$ for each partition contributing.

4. Model symmetries and data symmetries

According to the formula at the end of the previous subsection, the $N = 5$ sequence (i.e., slowness model) in equation (9), which has 7 partitions, each of which can be re-ordered in 5! different ways, produces a total of 7×5! = 840 different realizations that preserve the sum (i.e., traveltime). This in some sense helps us to quantify the non-uniqueness of a travel time measurement associated with a ray crossing 5 discrete slowness model elements: if you measure a certain travel time, you have constrained the model to be one of 840 possibilities.

However, a case can be made that we have over-counted. Notice, for instance, that in getting the number 840, we have said that there are 120 ways of re-ordering each row of equation (9), including the bottom row. But in each re-ordering of the bottom row we are exchanging 1s with other 1s, always producing the same sequence $\{1, 1, 1, 1, 1\}$. Do these reproductions of the same sequence count? Or, should we be asserting that there is really

Partition	n_0	n_1	n_2	n_3	n_4	n_5	$N!/n_0!n_1!n_2!n_3!n_4!n_5!$
$\{5,0,0,0,0\}$	4	0	0	0	0	1	5
$\{4,1,0,0,0\}$	3	1	0	0	1	0	20
$\{3,2,0,0,0\}$	3	0	1	1	0	0	20
$\{3,1,1,0,0\}$	2	2	0	1	0	0	30
$\{2,2,1,0,0\}$	2	1	2	0	0	0	30
$\{2,1,1,1,0\}$	1	3	1	0	0	0	20
$\{1,1,1,1,1\}$	0	5	0	0	0	0	1

Table 1. Re-orderings within partitions: $N = 5$ case.

only one ordering of this row, since there is no change in the sequence when elements are exchanged, and doing our counting that way?

There appear to be two slightly different types of change we can make to a model such that the travel time is preserved. One is sort of trivial, in which the travel time is preserved because the model itself is preserved. If we switch the order of the elements of $\{1,1,1,1,1\}$ the *model* is symmetric under this operation; *a fortiori*, so are the data. The other is when the model undergoes a change, in either the values of slowness or the organization of the slownesses. We will have to make a distinction between these situations, because how we count models will be different.

5. Excluding model symmetries

Let us work out the right formula for the situation in which model-symmetries are excluded from the counting — i.e., that there is only one way of organizing $\{1,1,1,1,1\}$, but quite a few for $\{3,1,1,0,0\}$. We can figure this out by looking back at the example in equation (13), and imagining that, in the sequences in equation (13), the a, b, c, and d were not all distinct. For a specific example, suppose that $d = b$. If we go through equation (13), sequence by sequence, and replace each d with a b, we find that each of the 24 elements now has a counterpart that is equal to it. For instance, the top left sequence becomes equal to the far right sequence second from the bottom. In this case, the number of distinct sequences has evidently been cut in half: $4!/2! = 12$. If we then found out that a also was equal to b and d, and replaced each a with a b, only 4 different distinct sequences would be left[†]: $4!/6 = 4!/3! = 4$. If there are two different 'clumps' of identical elements, say $a = c$ and $b = d$, instead the number of distinct elements we come up with is $4!/(2! \times 2!) = 6$. Gradually the pattern emerges: given N elements of a sequence, with n_0 of these sharing one value (0 being a convenient one), n_1 of them sharing another value, and so on up to n_N, the total number of distinct sequences is $N!/(n_0!n_1!n_2!...n_N!)$, or more compactly $N!/\Pi_j n_j!$. Notice that each partition of N has a different set of n_i values. For instance, each partition of $N = 5$ has the n_i's given in Table 1. Consequently the number

[†]That is, the 4 different places a 'c' can be placed amongst 3 'b's.

of *unique* slowness models which give rise to the same traveltime is

$$n_\tau = \sum_{i=1}^{p(N)} \frac{N!}{\Pi_{j=0}^{N} n_j(i)!}. \tag{16}$$

However moving from this formula (in which we assume we have in advance all the $n_j(i)$) to a straightforward enumeration is not straightforward.

Stars and bars

The above is where I brought the counting to, and it hits a bit of a wall. Fortunately, this problem as enumerated in equation (16) can be solved through a different combinatorial approach which is referred to as the "stars and bars" problem, which Michael Lamoureux showed me; and in fact the problem as we have set it out can be answered very simply:

$$n_\tau = \frac{(2N-1)!}{N!(N-1)!}. \tag{17}$$

CONCLUSIONS

To complete the process of counting, the way it has been set out here, we need to answer a question which is tough to answer the way the problem is come at here – what is $n_j(i)$ in general? We have a formula for the number of partitions of N there are, so if we decide to allow model symmetries, then we have an exact count of the number of models and the non-uniqueness of the travel time, equation (14). However, if we disallow model symmetries, which we almost certainly should, we hit a wall analytically using the partitions approach. Fortunately, a different combinatorial approach gives us a straightforward answer.

ACKNOWLEDGEMENTS

Thanks to Michael Lamoureux for pointing me to the Stars and Bars problem. We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13, and in part thanks to the Canada First Research Excellence Fund.