Soduko: Minlex Form and Chaining
by
Michael Deverin

### Minlex Form

There is a set of operations which can be performed on Soduko puzzles (both completed and unfinished), which keep the essential nature of the puzzle unchanged. Clues can be relabeled using any permutation of digits 1 - 9; rows may be interchanged within a band; columns may be interchanged within a chute; bands may be swapped; chutes may be swapped; the grid may be transposed. Soduko enthusiasts with programming skills sometimes wish to transform grids to a standard form. The normal reason for doing this is to prevent multiple versions of the *essentially* same puzzle from being saved separately.

One of the canonical forms for this purpose is minimum lexicographic form ("minlex" form). The result of the transformation is the smallest obtainable 81-digit number, when the puzzle is viewed as the concatenation of all the rows in normal order. The digit zero represents any unknown cell. The remainder of this paper describes a technique called chaining which makes conversion of **solution grids** to minlex form very fast.

### The First Two Rows

If row-1 is fixed to be '123456789', there are exactly 12096 (= 56x6x6x6) ways to fill row-2. Using brute force technique, each of these two-row pairs were transformed to minlex form. The procedure produced only 15 different results. In fact, using interchange and relabeling operations, any pair of rows within a band can be transformed to one of these fifteen choices. See "Table 'A': Two-Row Minlex Form". Note that entries are in ascending order. The **rank** of an entry is its position in the table.

For some ranks, there are multiple interchange/relabeling combinations which lead to the same minlex result. See Table 'A', under heading '#ways'. The following example shows how the same two rows can be transformed in three different ways to reach the same result.

 123 456 789 123 456 789 123 456 789

 Row-1: 469 172 853 469 172 853 469 172 853

 Row-2: 873 965 421 873 965 421 873 965 421

 Key #1 Key #2 Key #3

 132 798 546 546 231 897 897 645 132

 Row-1a: 496 835 712 712 694 538 538 217 496 (reordered)

 Row-2a: 837 412 695 695 738 214 214 596 837

 Row-1b: 123 456 789 123 456 789 123 456 789 (relabeled)

 Row-2b: 457 189 326 457 189 326 457 189 326

The **rank** of '457 189 326' is six, since it is the sixth entry in Table 'A'. It may come as no surprise, that if we examine the same two rows in flip-flop order, we can also find three permutations which again produce the same minlex result.

 123 456 789 123 456 789 123 456 789

 Row-2: 873 965 421 873 965 421 873 965 421

 Row-1: 469 172 853 469 172 853 469 172 853

 Key #4 Key #5 Key #6

 645 897 231 798 132 645 231 546 798

 Row-2c: 596 214 738 412 837 596 738 695 412 (reordered)

 Row-1c: 217 538 694 835 496 217 694 712 835

 Row-2d: 123 456 789 123 456 789 123 456 789 (relabeled)

 Row-1d: 457 189 326 457 189 326 457 189 326

What may be surprising is that second set of keys is the same as the first set -- but in reverse order! Check it out. This **reverse order** **phenomenon** is true for only ***some*** ranks, namely ranks 1, 3, 4, 6, 9, 12 and 13. However, for ***every*** rank, there is a permutation which enables direct conversion to the corresponding set of **dual** keys. Thus, if we are able to produce one key set, the dual key set is immediately available. See Table 'A', under heading 'Dual Key Permutation.'

### Two-Row Chaining used to Determine Rank

Chaining is a fast way to identify which row pairs are candidates to be the first two rows of the standard form matrix. Chaining will be described by example:

 Chute: a a a b b b c c c

 Row-1: 4 6 9 1 7 2 8 5 3

 Row-2: 8 7 3 9 6 5 4 2 1

 Cycles: (4 8)(6 7)(9 3 1)(2 5)

 ABC #1: (a c)(a b)(a c b)(b c)

The "head" of the chain is the vertical pair of digits in column-1. We then take the row-2 digit just found and find its column in row-1, and then record the row-2 digit in that same column as the next member of the cycle. When the initial digit of the cycle is found in row-2, the cycle ends, and we start a new cycle.

Chaining is just a way to produce the cycle structure for the two rows. Looking at the cycle as "chute" identifiers changes the focus to something which may be common to a large set of row pairs. The initial 'abc' chute designation was quite arbitrary. The full set:

 All Six 'abc' Choices Reordered / Alphabetical

 ABC #1: (a c)(a b)(a c b)(b c) (a b)(a c)(b c)(a c b)

 ABC #2: (a b)(a c)(a b c)(c b) (a b)(a c)(b c)(a b c) \*

 ABC #3: (b a)(b c)(b a c)(c a) (a b)(a c)(b c)(a c b)

 ABC #4: (b c)(b a)(b c a)(a c) (a b)(a c)(b c)(a b c) \*

 ABC #5: (c b)(c a)(c b a)(a b) (a b)(a c)(b c)(a c b)

 ABC #6: (c a)(c b)(c a b)(b a) (a b)(a c)(b c)(a b c) \*

In the last column, the cycles are presented, first by cycle length, and then in alphabetic order within cycle. Finally, from the re-ordered patterns, we choose the one with the earliest dictionary order as the **cycle standard form** (the ones marked with an '\*').

Here is the punch line:
 **There are exactly 15 values for cycle standard form and each value
 corresponds to exactly one of the two-row minlex forms.**

### Computational Short Cuts

Wait, you say! Isn't the cycle-standard-form procedure expensive to perform? The answer is "no". The nice thing is we can short-cut the procedure and stop after ABC #1 is computed. We just keep a bigger table of items which map back to the 15 minimal forms. Basically, we: 1) compute a value representative of ABC#1; 2) use a binary search to find that value in a pre-computed table (my table is 224 long); and 3) return the "rank" which corresponds to the value.

To ease subsequent computations, compute the puzzle transpose early, so that chutes just can be treated as three additional bands. To find candidates for the first two rows in the minlex matrix, we need to "rank" pairs of rows within each of our *six* bands. Only three routine calls per band need be made, as the rank for row pairs in opposite order is the same.

### Survey Results

In a survey of one-million random grids, 18 routine calls per grid were made to find the lowest ranking row pair. In just over 70% of the cases, the lowest ranking pair was of rank = 4. This is especially fortuitous, since there is only one way to convert the row pair to minlex form. As a weighted average, including the dual key set, there are 10.67 interchange/relabelings for the first two rows. (In the example, there were 6 ways that '457 189 326' could be generated. See "Table 'B': Solution Grid Lowest Rank Survey".

### Computing Keys from a Pair of Rows

Given two rows that are candidates for the first two rows of the minlex matrix, the set of transformation keys can be computed by mimicking the known structure of the target minlex form. This procedure is quite detailed and beyond the scope of this paper. However, once the key set is in hand, the dual key set is immediately available.

### After the First Two Rows

Suppose extend our list of two-row putative answers to include row-3. We know that it is the only row remaining in the band. We also know the full column key and the clue substitution permutation. We know everything needed to compute a 9-digit "score". Only three-row putative answers with lowest score need be kept. Normally, there will only a few answers in the three-row list.

The row order for minlex form rows 4 - 9 is easily computed. Simply use the clue substitution map to relabel the remaining column-1 digits in original grid. The first row of the second band in the minlex matrix will begin with a relabeled digit '2'.

### Time Tests

In timing tests on my home computer (Dell Studio 540, Intel Core, 2 Quad CPU, Q8200 @ 2.33 GHz) full random soduko grids were converted to minlex form at a rate of just over 50,000 per second.

Michael A. Deverin
August 1, 2011

|  |
| --- |
|  Table 'A': Two-Row Minlex Form |
| Rank | Row-1 = '123456789'Row-2 Minlex Form | Frequency | #Ways | Cycle Standard Form | Dual KeyPermutation |
| 1 | 456 789 123 |  72 |  18 | (abc)(abc)(abc) | 9,8,7,6,5,4,3,2,1 |
| 2 | 456 789 132 |  216 |  6 | (abc)(abcabc) | 1,2,3,7,9,8,4,6,5 |
| 3 | 456 789 231 |  144 |  9 | (abcabcabc) | 9,8,7,6,5,4,3,2,1 |
| 4 | 457 189 236 |  1296 |  1 | (ab)(ac)(abcbc) | 9,8,7,6,5,4,3,2,1 |
| 5 | 457 189 263 |  1296 |  1 | (ab)(abcbcac) | 4,5,6,1,2,3,8,7,9 |
| 6 | 457 189 326 |  432 |  3 | (ab)(ac)(bc)(abc) | 9,8,7,6,5,4,3,2,1 |
| 7 | 457 189 623 |  1296 |  1 | (ab)(abc)(acbc) | 4,5,6,1,2,3,7,8,9 |
| 8 | 457 189 632 |  1296 |  1 | (ab)(abcacbc) | 4,5,6,1,2,3,7,9,8 |
| 9 | 457 289 163 |  432 |  3 | (ababcbcac) | 9,8,7,6,5,4,3,2,1 |
| 10 | 457 289 613 |  1296 |  1 | (abac)(abcbc) | 5,4,6,2,1,3,7,8,9 |
| 11 | 457 289 631 |  1296 |  1 | (ababcacbc) | 5,4,6,2,1,3,7,9,8 |
| 12 | 457 389 612 |  1296 |  1 | (abacbacbc) | 9,8,7,6,5,4,3,2,1 |
| 13 | 457 389 621 |  1296 |  1 | (abc)(abacbc) | 9,8,7,6,5,4,3,2,1 |
| 14 | 457 893 612 |  216 |  6 | (abc)(abc)(acb) | 1,2,3,8,9,7,6,4,5 |
| 15 | 457 893 621 |  216 |  6 | (abc)(acbacb) | 1,2,3,9,8,7,6,5,4 |
|  |  |  12096 |  |  |  |

|  |
| --- |
| Table 'B': Solution Grid Lowest Rank SurveyFor each of 1,000,000 random solution grids, 18 routine calls were made to find the lowestranking row/column minlex pair. For 89290 grids (Row Sum #1), the minimum rank was 1.Of those the row/column pair was unique in 74566 cases, and occurred twice 11031 times. |
| #Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Row Sum |
| 1 | 74566  | 11031  | 3401  | 246  | 34  | 11  |  1 | -- | -- | -- | 89290 |
| 2 | -- | 90415 | -- | 4269 | -- | 116 | -- | 4 | -- | -- | 94804 |
| 3 | -- | -- | 12509 | -- | -- | 72 | -- | -- | 1 | -- | 12582 |
| 4 | 216215 | 229724 | 154633 | 72209 | 25265 | 6877 | 1378 | 236 | 33 | 2 | 706572 |
| 5 | 22760 | 27073 | 20201 | 10979 | 4372 | 1294 | 328 | 52 | 9 | -- | 87068 |
| 6 | 3522 | 1437 | 385 | 70 | 13 | -- | -- | -- | -- | -- | 5427 |
| 7 | 755 | 1054 | 994 | 664 | 343 | 123 | 35 | 12 | 1 | 1 | 3982 |
| 8 | 32 | 74 | 69 | 50 | 35 | 9 | -- | 1 | -- | -- | 270 |
| 9 | 1 | 2 | -- | -- | -- | -- | -- | -- | -- | -- | 3 |
| 10 | 1 | -- | 1 | -- | -- | -- | -- | -- | -- | -- | 2 |
|  | 317852 | 360810 | 192193 | 88487 | 30062 | 8502 | 1742 | 305 | 44 | 3 | 1000000 |