

Solution to four-color problem

Hamid V. Ansari

Department of Physics, Isfahan University, Isfahan, IRAN
Personal address: No. 16, Salman-Farsi Lane, Zeinabieh Street, Isfahan,
Postal Code 81986, IRAN

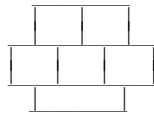
September 19, 2007

Abstract

To color a given map we first find its related map with the most mutual adjacencies and color it by only four colors, then we trace back.

1 Introduction

It can be proven easily that on a planar map there can be no more than four mutually adjacent regions. Since every two regions require two different colors to be distinguished, it seems that we can conclude that we won't need more than four colors to color the regions of the map distinctly. That we need no more than four colors to color any planar map needs to be proven, because for instance we have a map consisting of no more than three mutually adjacent regions but requiring at least four colors to be colored; see the following figure.



2 The proof

First we color all the regions and the background on the given map by the color 1: see Fig. 1. Then we choose a region and color it by 2: see Fig. 2. If no region is adjacent to this region or a region surrounds it, we can use its color (2) to color other regions; in this state we say we ignore this region. But if a region is adjacent to it we color it by 3: see Fig. 3.

Now we can ignore these regions if no region is adjacent to any one of them or a region surrounds both of them. If a region surrounds only one

of them we can ignore the surrounded region and the situation is as if we have again two adjacent regions. Thus, we recognize two states:

1. A region is adjacent to only one of these region.
2. A region is adjacent to both of these regions.

In the first state we have the general state shown in Fig. 4. (Of course Fig. 4' is equivalent to Fig. 4, and then we don't discuss about it.)

We suppose that the line between the two points "a" and "b" is not crossed by any border of any other region; in fact we choose the first region adjacent to 2 and most near to the right 3 as the left 3 (Fig. 4). In such a case we can slip the points "a" and "b" toward each other along the line ab and continue to slip one of them along a part of the border of the other region to obtain Fig. 5 in which we have changed the color of one of the two previously 3-colored regions.

Instead of coloring Fig. 4 we can try to color mutually adjacent regions of Fig. 5 and then use the same coloring for Fig. 4. That we suppose that there is no cross points on ab and that brief displacement of the probable cross points caused by the borders of other regions on the borders of 3 and 4 (without altering their order) is possible proves that each form of coloring applicable to Fig. 5 can be used for Fig. 4 too. About how Fig. 5 is colorable we speak soon.

Now we consider the second state. There is a region, colored by 4 in Fig. 6, which is adjacent to both of the regions 2 and 3. (Fig. 5 is of this kind.) Now, if no region is adjacent to these regions or a region surrounds all of these regions we can ignore them. If a region is adjacent to only one of these regions the general state shown in Fig. 7 will be obtained. In this figure we have chosen that region (adjacent to 3) that causes the line cd not to be crossed by any border of any region. In this state in a manner similar to what used in obtaining Fig. 5 from Fig. 4 we can obtain Fig. 8 (with a similar reasoning). If we can color Fig. 8, certainly its coloring form will be applicable to Fig. 7. Thus, we should see if Fig. 8 is colorable, and this is the next subject.

Now if a region is adjacent to only two regions of Fig. 6, then the general state shown in Fig. 9 will be obtained in which the new region has been colored by 2.

Now we want to see, continuing our manner, if we can change the subject of coloring such a set into the subject of coloring a set of mutually adjacent regions. It is possible: Let's construct Fig. 10 from Fig. 9. We have only extended the top region 2 over 4 reaching the under region 2, the only condition being that all the cross points of the borders of other probable regions along ef has been transferred to the partial border EF (without any change in their order).

It's evident that the coloring performed on Fig. 10 is not valid and then we change it into Fig. 11 in which only the colors of the extended top region and the background have been changed.

Now we want to return to our original figure, Fig. 9. After that eventually (following any procedure) all the regions, that are adjacent to 1 along EF, have been colored, certainly none of them will have the color 1. This says to us that when returning from Fig. 11 to Fig. 9, if after coincidence of EF with ef we only change the color of the central region (in Fig. 11) from 4 to 1, no problem will occur; in this state it will be also necessary to change the color of the remaining top region from 1 to 2 which is in fact its original color (in Fig. 9). In this manner we reach eventually to Fig. 12 which is equivalent to Fig. 9 (because in Fig. 9 we are free to change the colors 4 and 1 to each other):

And eventually, we have no other choice to find its related set of mutually adjacent regions. As it is seen, our lemma in proving the four-color theorem for a given map is providing its related map which has the most mutual adjacencies. Every form of coloring applicable to this related map will be suitably applicable to the original map when tracing back. Therefore, if we can prove the four-color theorem for a map with the most mutual adjacencies, this theorem will have been proven for any map.

The proof of the 4CT for a map with the most mutual adjacencies:

Choose a region and color it by 1. Choose a region adjacent to it and color it by 2. Choose a region adjacent to both of these colored regions and color it by 3. Choose a fourth region adjacent to all of these three regions and color it by 4. This region causes one of the regions 1, 2 and 3 to become an inside region capable of being ignored in coloring other regions; thus we use the color of this inside region in coloring the next region. This next region should be chosen adjacent to 4 and those two regions out of 1, 2 and 3 which have not become inside regions. This next region again surrounds one of the colored regions which have not become inside regions yet, and we can continue the same procedure to color all the regions using only 4 colors.

The general figure of a map with the most mutual adjacencies is as shown in Fig. 13 or Fig. 14 or a combination of these.

As an example in the following we see a map with the most mutually adjacent regions but with a lower order (because in it one region does not follow the rules of the game, and probably this is why eventually we will encounter two choices when coloring it directly through the above-mentioned algorithm presented for coloring a map with the most mutual adjacencies (without previous using of the presented lemma).)

It may be in the form as shown in Fig. 15. And its equivalent form is as shown in Fig. 16. (As it is seen, we have incompatibility in the region “=”.)

In practice there is no difference between the lemma that changes Fig. 4 to Fig. 5 and the lemma that changes Fig. 7 to Fig. 8 and the lemma that changes Fig. 9 to Fig. 10, because in the second and third ones we do in fact the same act as done in the first one.

As an exercise let's color the map shown in Fig. 17 according to the proof and using its lemma to obtain Fig. 18 and then Fig. 19. This is now a map with the most mutual adjacencies and can be colored using the presented algorithm to obtain for instance $a = 1, d = 2, c = 3, b = 4, e = 1$ and $f = 4$, which are applicable to the original map.

It's evident that every change in the boundaries of a map with the most mutual adjacencies in such a manner that creates a necessity for a new color is simultaneous with removal of the necessity for an existent color. Naturally (of course not probably directly but with several color displacing depending on the complexity of the map), the color which its necessity has been removed will be used as the new necessary color. Thus, we can conclude intuitively that every change in the boundaries of a map with the most mutual adjacencies causes no necessity to any fifth color.

It's evident (or can be proven as a theorem) that every configuration of maps can be obtained by proper changes in the boundaries of proper maps with the most mutual adjacencies. If so, considering the conclusion of the previous paragraph we infer that there is no map needing more than four colors.

And we can follow a constant (but not straightforward) method for coloring all the planar maps: we should create related map with the most mutual adjacencies and color it by four colors and then return step by step in each of which doing all the color displacements initially caused due to construction of the map with the most mutual adjacencies using the stated lemmas and we should consider all the effects of these color displacements on all the patches of the map. (Of course this act in complex maps is difficult but is a constant method. I think proof for its truth and applicability is parallel with any proof for the above-mentioned theorem (ie this fact that suitable changes in the boundaries of maps with the most mutual adjacencies will create every wanted configuration of any map)).

As an example suppose that the entire map looks like Fig. 20. To construct its related map with the most mutual adjacencies we apply two transformations:

- 1: We extend the boundary of $2'$ over 4 to reach to 2 .
- 2: We extend the boundary of E over $2'$ to reach to A .

After these transformations the related map with the most mutual adjacencies will be as shown in Fig. 21.

Using the presented algorithm we can easily color this map with four colors. The colors that these regions get in this manner won't be applicable to the original map unless we consider displacement of colors when we reverse the applied transformations. If one is patient can consider these displacements (and esp their effects on other regions) until the final coloring (correct for the original map) is obtained. Anyway, since reversing each of the above transformations creates the necessity for a new color simultaneous with removing the necessity for an existent color, it is evident that reversing these transformations won't create any necessity for a new

color but only, during the reversing, the colors of the regions will be displaced suitably until the final (and in fact original) map will be obtained (with the correct coloring with at most four colors).

Let's see the process of reversing in this example: Let our four colors be a, b, c and d .

Beginning from the recent map (with the most mutual adjacencies):

	a	b	c	d
0:	2	3	4	2'
	E	B	A	D
		C	1	
1:	2	3	4	D
	E	B	A	
	2'	C	1	
2:	2	3	A	D
	E	B	1	4
	2'	C		
3:	2	3	A	D
	2'	B	1	4
		C		E
4:	2	3	A	4
	2'	B	1	E
	D	C		

This coloring is applicable to the map shown in Fig. 22.

And we continue:

	a	b	c	d
0':	2	3	A	4
	2'	B	1	E
	D	C		
1':	2	3	A	4
	2'	B	1	
	D	C	E	
2':	2	3	A	4
	2'	B	1	C
	D	E		
3':	2	3	A	4
	2'	B	E	C
	D	1		

And this coloring is applicable to the (original) map as shown in Fig. 23.

1

Fig. 1

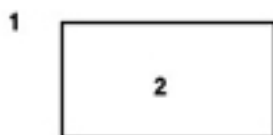


Fig. 2



Fig. 3

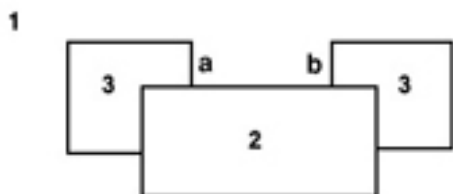


Fig. 4

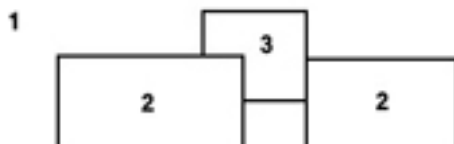


Fig. 4'

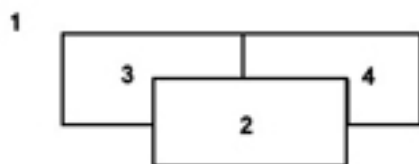


Fig. 5

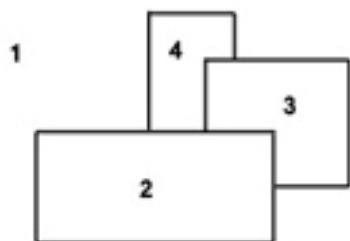


Fig. 6

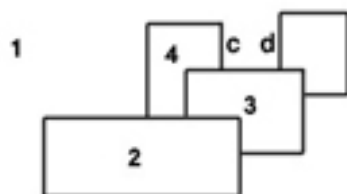


Fig. 7

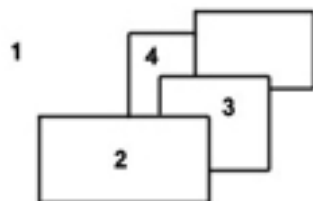


Fig. 8

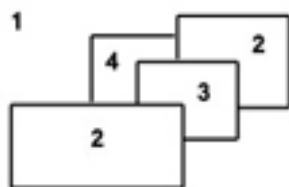


Fig. 9

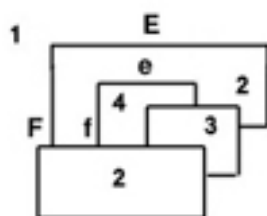


Fig. 10

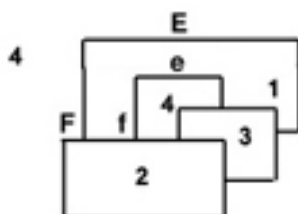


Fig. 11

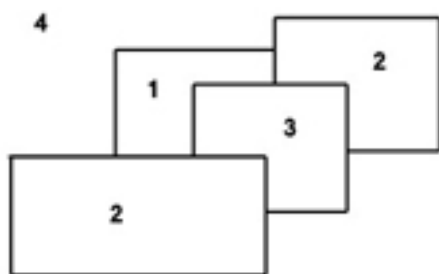


Fig. 12

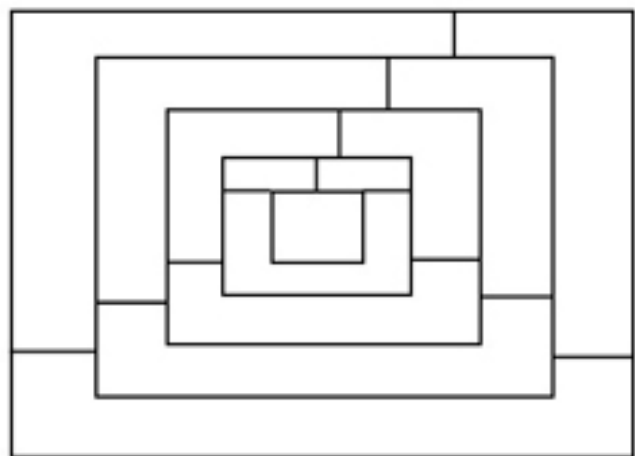


Fig. 13

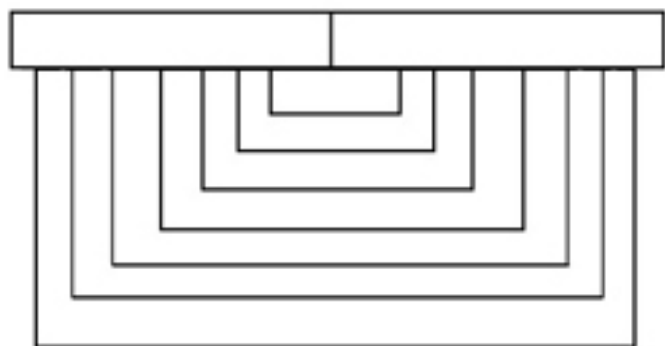


Fig. 14

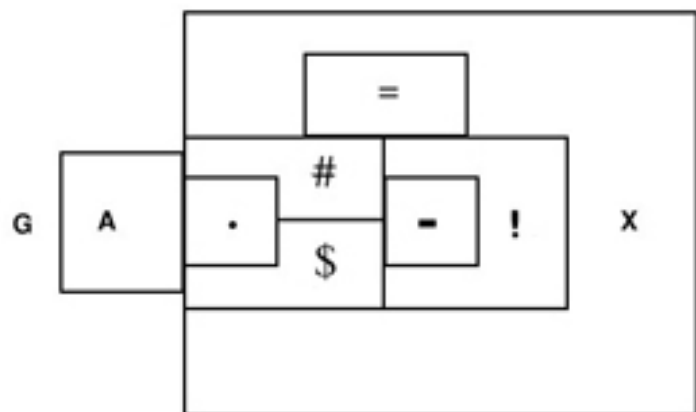


Fig. 15

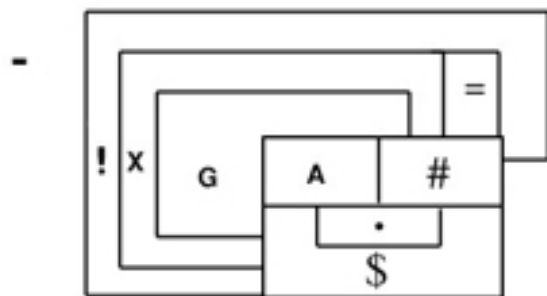


Fig. 16

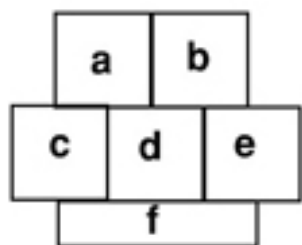


Fig. 17

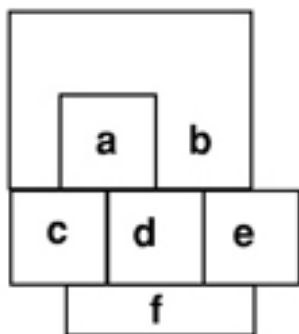


Fig. 18

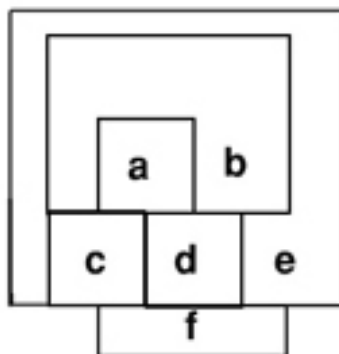


Fig. 19

1

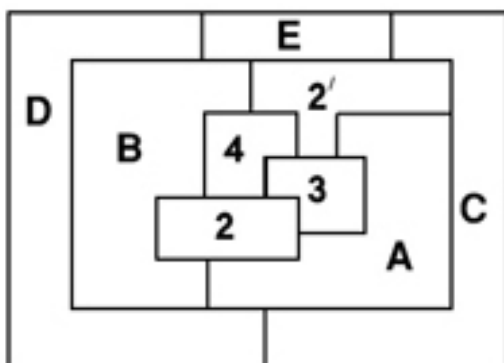


Fig. 20

1

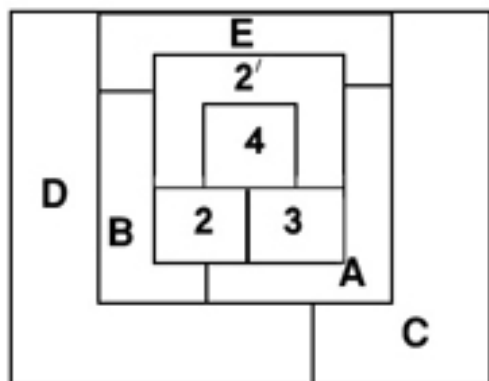


Fig. 21

1

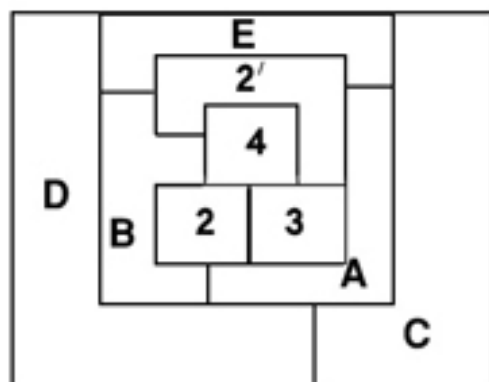


Fig. 22

1

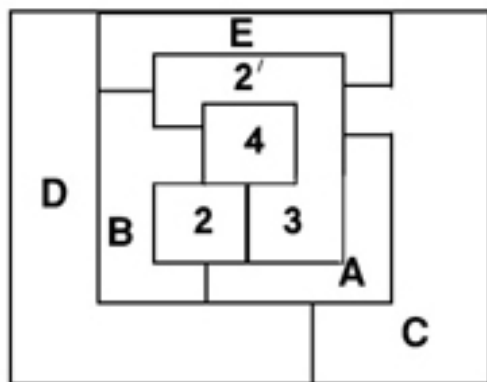


Fig. 23