Theorem 1.49: van Kampen Theorem

We fix the following situation:

- (X, x_0) a pointed space,
- $\{\mathrm{A}_\alpha\}_\alpha$ a covering of X by **path-connected open** subsets such that $x_0 \in \mathrm{A}_\alpha$ **for all** α.

In the following all fundamental groups are with respect to the base point x_0. This situation gives us a number of maps and group homomorphisms:

- The inclusion $i_\alpha : \mathrm{A}_\alpha \to \mathrm{X}$ induces a homomorphism $\varphi_\alpha : \pi_1(\mathrm{A}_\alpha) \to \pi_1(\mathrm{X})$ (this was Proposition 1.34).
- This induces a homomorphism $\varphi : *_\alpha \pi_1(\mathrm{A}_\alpha) \to \pi_1(\mathrm{X})$ (this was Lemma 1.47).
- The inclusion $i_{\alpha\beta} : \mathrm{A}_\alpha \cap \mathrm{A}_\beta \to \mathrm{A}_\alpha$ induces a homomorphism $\varphi_{\alpha\beta} : \pi_1(\mathrm{A}_\alpha \cap \mathrm{A}_\beta) \to \pi_1(\mathrm{A}_\alpha)$ (this was Proposition 1.34).

The van Kampen theorem gives us conditions on when φ is surjective and under which additional condition one can deduce the kernel of φ.

Theorem 1.49 (van Kampen Theorem)**.** *Let* (X, x_0), $\{\mathrm{A}_\alpha\}$, φ, *and* $\varphi_{\alpha\beta}$ *defined as above.*

(a) If each $\mathrm{A}_\alpha \cap \mathrm{A}_\beta$ *is path-connected, then* φ *is surjective.*

(b) If each $\mathrm{A}_\alpha \cap \mathrm{A}_\beta \cap \mathrm{A}_\gamma$ *is path-connected, then the normal subgroup* N *generated by all* $\varphi_{\alpha\beta}(\omega)\varphi_{\beta\alpha}(\omega)^{-1}$ *for* $\omega \in \pi_1(\mathrm{A}_\alpha \cap \mathrm{A}_\beta)$ *and all* α *and* β*, is the kernel of* φ*. Especially* φ *induces an isomorphism of groups*

$$\pi_1(\mathrm{X}) \cong (*_\alpha \pi_1(\mathrm{A}_\alpha))/N.$$

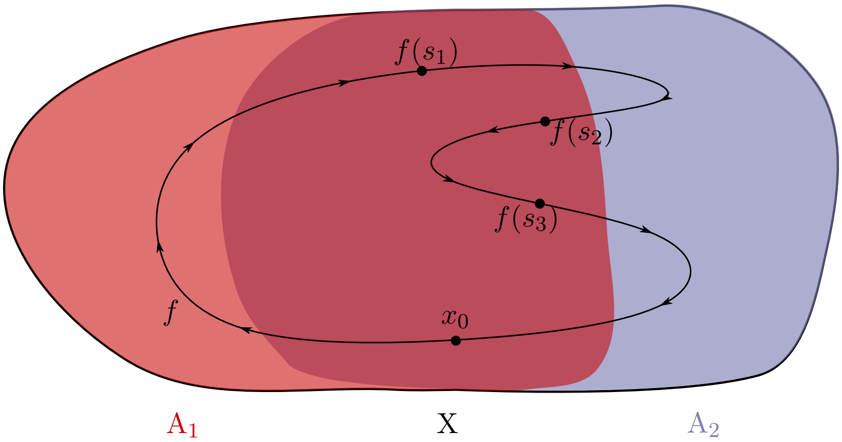
Remark: Note that the subgroup N is generated as a **normal** subgroup by elements of the form $\varphi_{\alpha\beta}(\omega)\varphi_{\beta\alpha}(\omega)^{-1}$. Thus elements of N are finite products of elements of the form $g\varphi_{\alpha\beta}(\omega)\varphi_{\beta\alpha}(\omega)^{-1}g^{-1}$ for $g \in *_\alpha \pi_1(\mathrm{A}_\alpha)$.

We will first prove part (a) and afterwards formalize the result before we proceed with part (b). Note that it is easy to check that elements of the form $\varphi_{\alpha\beta}(\omega)\varphi_{\beta\alpha}(\omega)^{-1}$ are always contained in the kernel of φ. This does not need any extra assumptions on the intersections of the A_α. Hence in every case N is contained in the kernel. We will only need to prove the opposite inclusion.

Theorem 1.49 (van Kampen Theorem). *(a) If each $\mathrm{A}_\alpha \cap \mathrm{A}_\beta$ is path-connected, then φ is surjective.*

Proof. of part (a) Fix $f : \mathrm{I} \to \mathrm{X}$ a loop at x_0. Since f is continuous: for every $s \in \mathrm{I}$ and α with $f(s) \in \mathrm{A}_\alpha$, there is an open interval $s \in \mathrm{I}_{s,\alpha} \subset \mathrm{I}$ with $f(\mathrm{I}_{s,\alpha}) \subset \mathrm{A}_\alpha$. Since I is compact: finitely many of these open intervals cover I. Thus we can choose $0 = s_0 < s_1 < \ldots < s_m = 1$ such that $f([s_{i-1}, s_i]) \subset \mathrm{A}_{\alpha(i)}$ for some $\alpha(i)$. In an example pictures this would look as follows:

The picture to the right shows a path f with choices s_1, s_2 and s_3. The parts of f between two consecutive points are contained entirely in one of the two open subsets A_1 or A_2. This is not unique and we do not assume that it is minimal. One could delete the points s_2 and s_3 here and the property would still hold.

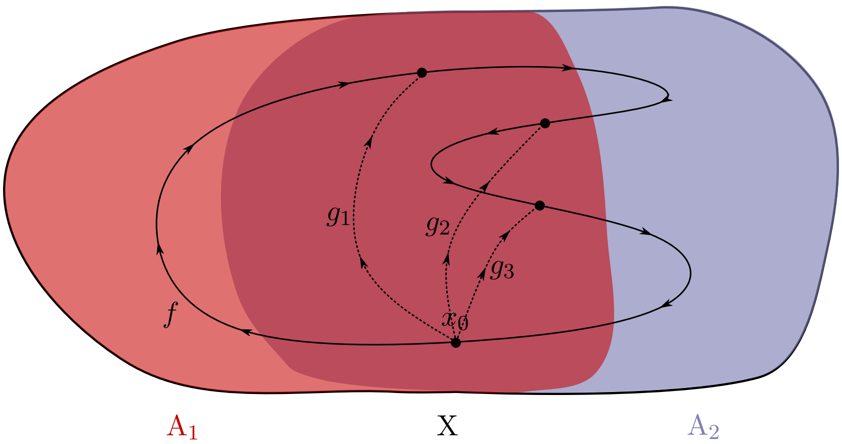


Define f_i to be f restricted to the interval $[s_{i-1}, s_i]$. Then f_i is a path from $f(s_{i-1})$ to $f(s_i)$ and $f \simeq f_1 \cdots f_m$. Then $f(s_i) \in \mathrm{A}_{\alpha(i)} \cap \mathrm{A}_{\alpha(i+1)}$, which is path connected by assumption, hence we can choose g_i a path in $\mathrm{A}_{\alpha(i)} \cap \mathrm{A}_{\alpha(i+1)}$ from x_0 to $f(s_i)$. In addition let g_0 and g_m both be the constant path at x_0. Then

$$f \simeq (g_0 \cdot f_1 \cdot \overline{g}_1) \cdot (g_1 \cdot f_2 \cdot \overline{g}_2) \cdots (g_{m-2} \cdot f_{m-1} \cdot \overline{g}_{m-1}) \cdot (g_{m-1} \cdot f_m \cdot \overline{g}_m),$$

with each $g_{i-1} \cdot f_i \cdot \overline{g}_i$ being a loop inside $\mathrm{A}_{\alpha(i)}$ at x_0. In the picture from above this could look as follows:

We omit the two constant paths in here and only indicate the three that are non-trivial here.



Hence $[g_{i-1} \cdot f_i \cdot \overline{g}_i] \in \mathrm{im}(\varphi_{\alpha(i)}) \subset \mathrm{im}(\varphi)$. Thus $[f] \in \mathrm{im}(\varphi)$ and φ is surjective. $\square$

Factorizations of $[f]$: Fix $[f] \in \pi_1(\mathrm{X})$. A **factorization** of $[f]$ is an element $[f_1]\cdots[f_m] \in *_\alpha \pi_1(\mathrm{A}_\alpha)$ such that

- f_i is a loop in $\mathrm{A}_{\alpha(i)}$ at x_0, hence $[f_i] \in \pi_1(\mathrm{A}_{\alpha(i)})$ and
- $[f] = [f_1]\cdots[f_m]$ in $\pi_1(\mathrm{X})$, hence $f \simeq f_1 \cdots f_m$.

By definition a factorization is thus just a preimage of $[f]$ under φ. In part (a) we actually constructed such a factorization for a loop f, hence they always exist if the assumption of (a) is satisfied.

Note also that the assumptions for part (b) are stronger than the ones for part (a). Hence if we assume that triple intersections are path-connected as in part (b), then part (a) always applies and φ is already surjective. Now we need to compare different factorizations.

Equivalence of factorizations Two factorizations are equivalent if one can be obtained from the other by finitely many modifications of the forms:

(1) $[f_1]\cdots[f_i]\cdots[f_m] \sim [f_1]\cdots[f_{i-1}][f_{i+1}]\cdots[f_m]$ if $[f_i] = \mathbb{1}_{\pi_1(\mathrm{A}_{\alpha(i)})}$,

(2) $[f_1]\cdots[f_i][f_{i+1}]\cdots[f_m] \sim [f_1]\cdots[f_i \cdot f_{i+1}]\cdots[f_m]$ if $\alpha(i) = \alpha(i+1)$,

(3) $[f_1]\cdots[f_i]\cdots[f_m] \sim [f_1]\cdots[f_i]\cdots[f_m]$, where on the left we view $[f_i]$ as a class in $\pi_1(\mathrm{A}_\alpha)$ and on the right as a class in $\pi_1(\mathrm{A}_\beta)$ if f_i is a loop inside $\mathrm{A}_\alpha \cap \mathrm{A}_\beta$.

The modifications of type (1) or (2) are the same that are used in the definition of the free product. Hence factorizations equivalent via modifications of only type (1) or (2) are equal in $*_\alpha \pi_1(\mathrm{A}_\alpha)$. For the modification of type (3) we need to multiply the left hand side from the right with

$$([f_{i+1}]\cdots[f_m])^{-1}\varphi_{\alpha\beta}[f_i]^{-1}\varphi_{\beta\alpha}[f_i]([f_{i+1}]\cdots[f_m]) \in N$$

and obtain the right hand side. But that means that two factorizations of $[f]$, i.e. preimages of $[f]$ under φ differ by multiplication of an element of N. Hence both sides are equal in $*_\alpha \pi_1(\mathrm{A}_\alpha)/N$.

Theorem 1.49 (van Kampen Theorem). *(b) If each $A_\alpha \cap A_\beta \cap A_\gamma$ is path-connected, then the normal subgroup N generated by all $\varphi_{\alpha\beta}(\omega)\varphi_{\beta\alpha}(\omega)^{-1}$ for $\omega \in \pi_1(A_\alpha \cap A_\beta)$ and all α and β, is the kernel of φ. Especially φ induces an isomorphism of groups*

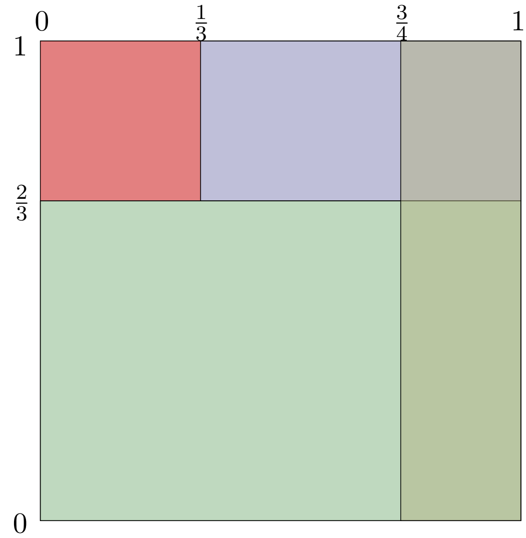
$$\pi_1(X) \cong *_\alpha \pi_1(A_\alpha)/N.$$

Proof. of part (b): From part (a) we know that φ is surjective. If we show that all factorizations of $[f]$ are equivalent then this means that φ factors through $*_\alpha \pi_1(A_\alpha)/N$. This implies that the kernel of φ is contained in N. Hence they are equal.
Let $[f_1]\cdots[f_m]$ and $[f'_1]\cdots[f'_n]$ be factorizations of $[f]$. Then $f_1\cdots f_m \simeq f'_1\cdots f'_n$ via a homotopy of paths $F : I\times I \to X$.
For $(s,t) \in I\times I$, choose a rectangle $(s,t) \in R_{s,t} = [a(s), b(s)] \times [c(t), d(t)] \subset I\times I$ with $F(R_{s,t}) \subset A_\alpha$ for some α, $a(s) < s < b(s)$, and $c(t) < t < d(t)$. This can be done since there is an open neighbourhood of (s,t) that maps into some A_α via F, since F is continuous and every open neighbourhood in $I\times I$ contains a closed rectangle. These $R_{s,t}$'s cover $I\times I$. By compactness of $I\times I$ a finite collection of these rectangles already covers $I\times I$. In an example this could look as follows:

Here we have four rectangles:

- The red shaded one: $[0, \frac{1}{3}] \times [\frac{2}{3}, 1]$
- The blue shaded one: $[\frac{1}{3}, 1] \times [\frac{2}{3}, 1]$
- The green shaded one: $[0, 1] \times [0, \frac{2}{3}]$
- The gray shaded one: $[\frac{3}{4}, 1] \times [0, 1]$

Note that these overlap, the gray one overlaps with both the blue and the green one in this case.



In this finite collection we order all the coordinates for the first interval to get a partition

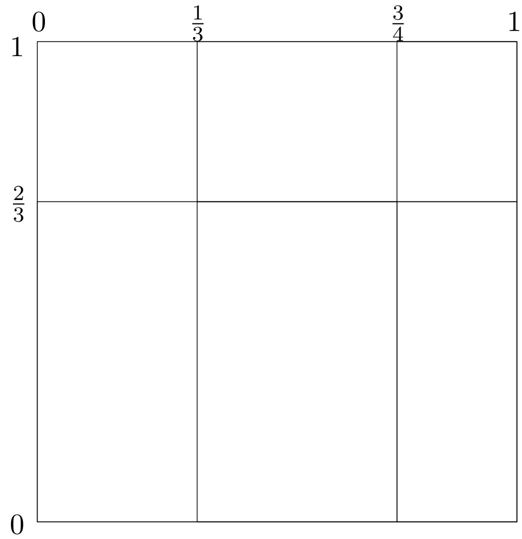
$$\underline{s} = (0 = s_0 < s_1 < \ldots < s_p = 1)$$

and all the coordinates of the second intervals to get a partition

$$\underline{t} = (0 = t_0 < t_1 < \ldots < t_q = 1).$$

Then, by construction, any $R_{i,j} = [s_{i-1}, s_i] \times [t_{j-1}, t_j]$ is mapped under F into some A_α. In our picture from above this would give the following:

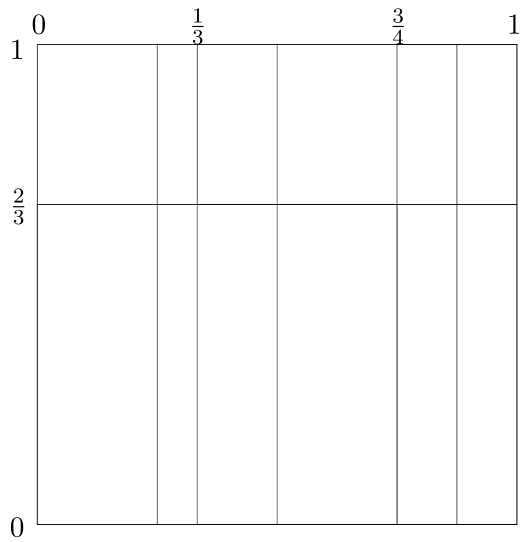
In this case we end up with a total of 6 rectangles that only overlap on their boundaries but not their interiors.



We will now need to refine and change these rectangles a bit to make them fit. This will be done in three steps.

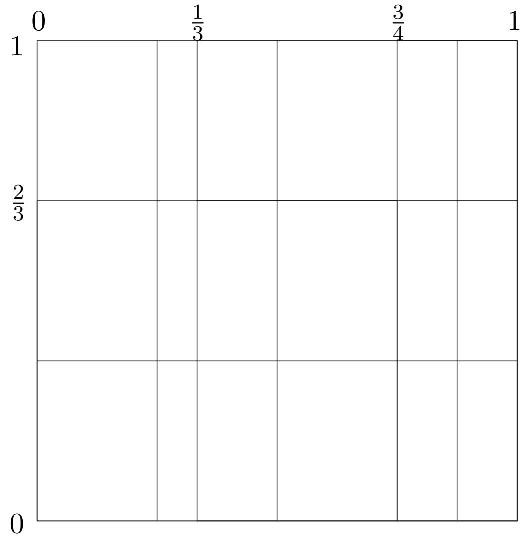
Change 1: Any choice of order of the multiplication for $f_1 \cdots f_m$ gives us a subdivision of I into m smaller intervals where on the i-th interval $f_1 \cdots f_m$ is given by f_i. For example $f_1 \cdot f_2$ has the two intervals $[0, 1/2]$ and $[1/2, 1]$. On the first interval the product is defined to be the path f_1 (up to reparametrization) and on the second interval the product is equal to the path f_2.
We make a choice of order of the multiplication for both factorizations we are given, $f_1 \cdots f_m$ and $f'_1 \cdots f'_n$, and assume that the coordinates $\underline{s}$ contain the coordinates for both of these subdivisions. This just makes the subdivision finer, hence it is still true that every rectangle gets mapped into some A_α via F. This is very easy to see in the picture:

Assume that we have the homotopy F from $(f_1 \cdot f_2) \cdot f_3$ to $f'_1 \cdot (f'_2 \cdot (f'_3 \cdot f'_4))$, with the two choices of the order of multiplication given by a choice of brackets. Then $(f_1 \cdot f_2) \cdot f_3$ is defined by using the intervals $[0, \frac{1}{4}]$, $[\frac{1}{4}, \frac{1}{2}]$, and $[\frac{1}{2}, 1]$. One the other hand $f'_1 \cdot (f'_2 \cdot (f'_3 \cdot f'_4))$ is defined by using the intervals $[0, \frac{1}{2}]$, $[\frac{1}{2}, \frac{3}{4}]$, $[\frac{3}{4}, \frac{7}{8}]$, and $[\frac{7}{8}, 1]$. Thus we add the coordinates $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{7}{8}$ into the partition $\underline{s}$ to obtain the rectangles to the right.



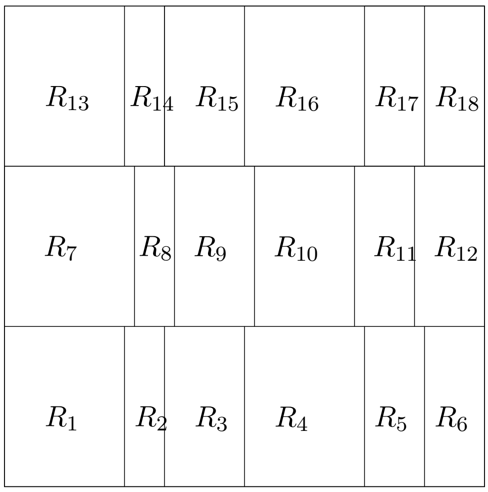
Change 2: We assume that there are at least 3 rows of rectangles, i.e. $q \geq 3$. This is just done by adding an additional coordinate into the partition $\underline{t}$ if $q < 3$. In pictures:

We can make any choice of an extra coordinate that is different from 0, $\frac{2}{3}$, and 1. To make it easy to draw, we thus just take $\frac{1}{3}$ and thus divide all the rectangles in the bottom row into 2 rectangles each.



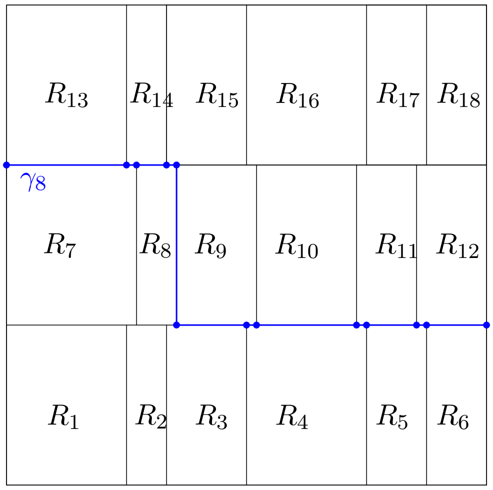
Change 3: This is the most important change. We assume that any point of $\mathrm{I} \times \mathrm{I}$ is contained in at most **three** rectangles. It is ovious that the only points in $\mathrm{I} \times \mathrm{I}$ where this can currently fail are interior corners of rectangles, that are each contained in four rectangles. For this we shift the rectangles in rows $2, \ldots, q-1$ horizontally so that it still holds that $F(R_{i,j}) \subset \mathrm{A}_\alpha$ for some α. This is possible because each rectangle has an open neighbourhood such that the open neighbourhood gets mapped into some A_α, since F is continuous. Thus we are allowed to slightly vary the coordinates of the second interval and the property still holds. The resulting rectangles we enumerate as $R_1, \ldots, R_{pq}$ from lower left to top right and denote by $\mathrm{A}_{\alpha(1)}, \ldots, \mathrm{A}_{\alpha(pq)}$ the open subsets they are mapped into. This is much easier to understand in a picture:

In the middle row of rectangles we change the coordinates for the second interval. As one can see this exactly means that any of the new corners are now only contained in at most three rectangles. One can choose any way of doing this. Like moving all of the rectangles a bit to the right, or as we show it here, some to the right and some to the left, or also completely random.



For a path γ from a point on the left edge of $\mathrm{I}\times\mathrm{I}$ to a point on the right edge, the composition $f_\gamma = F\circ\gamma$ is by construction a loop based at x_0. This holds since F was a homotopy of paths. We want to look at very specific choices for γ. Namely, for $0 < r < pq$ let γ_r be the shortest path in $\mathrm{I}\times\mathrm{I}$ that divides $\mathrm{I}\times\mathrm{I}$ into the area consisting of the rectangles $R_1,\dots,R_r$ and the area containing $R_{r+1},\dots,R_{pq}$. We set γ_0 as just the bottom edge of $\mathrm{I}\times\mathrm{I}$ and γ_{pq} as the top edge. In pictures this gives:

We choose here $r = 8$ and get the path γ_8.



Each γ_r is a product of straight line paths, the edges of the rectangles, with a number of distinguished points on these lines, the corners of the various rectangles. We will call the corners that such a path goes through its vertices and the line segments in between its edges.

In the above example we indicate the vertices on γ_8 by the blue dots, in total there are vertices $v_0,\dots,v_{12}$ on γ_8 with a total of twelve edges, i.e. line segments in between.

Fix $0 \le r < pq$. For a vertex v of γ_r we know by assumption that it is contained in at most three rectangles R_i, R_j and R_k. Choose a path g_v inside $\mathrm{A}_{\alpha(i)}\cap\mathrm{A}_{\alpha(j)}\cap\mathrm{A}_{\alpha(k)}$ from x_0 to $F(v)$. This is possible since we assume path-connectedness for these intersections. We denote by $v_0, v_1, \dots$ the vertices of γ_r and by e_l the edge between v_{l-1} and v_l. A factorization for $[f_{\gamma_r}]$ is then given by

$$[g_{v_0}\cdot f_{e_1}\cdot \overline{g}_{v_1}]\cdot[g_{v_1}\cdot f_{e_2}\cdot \overline{g}_{v_2}]\cdot[g_{v_2}\cdot f_{e_3}\cdot \overline{g}_{v_3}]\cdots,$$

where f_{e_l} is just F restricted to the edge e_l seen as a path. Note that whenever $F(v) = x_0$ for a vertex v we can just choose g_v to be the constant path, for example for $v = v_0$.
If an edge e_l is contained in two rectangles R_i and R_j, we can view $[g_{v_{l-1}}\cdot f_{e_l}\cdot \overline{g}_{v_l}]$ either as an element in the subgroup $\pi_1(\mathrm{A}_{\alpha(i)})$ or as an element in the subgroup $\pi_1(\mathrm{A}_{\alpha(j)})$. This gives two different factorizations. But they are equivalent by a single modification of type (3) by definition. Hence the choice of the rectangle is irrelevant up to equivalence of factorizations.
Comparing γ_r and γ_{r+1}, we see that they only differ by edges contained in R_{r+1}. By the previous argument we know that, up to equivalence, we can choose a factorization of γ_r where the factors for edges in R_{r+1} are already in $\pi_1(\mathrm{A}_{\alpha(r+1)})$. There is an obvious homotopy in R_{r+1} between γ_r and γ_{r+1}, hence the two factorizations are equivalent via type (2). Thus the factorization for f_{γ_r} and $f_{\gamma_{r+1}}$ are equivalent.
Inductively we now obtain that the factorizations we obtained for f_{γ_0} and $f_{\gamma_{pq}}$ are equivalent.

We are left to show that the factorization for f_{γ_0} is equivalent to $[f_1]\cdots[f_m]$ and that the factorization for $f_{\gamma_{pq}}$ is equivalent to $[f_1']\cdots[f_n']$. Both of these are done exactly the same, so we only give the argument for the f_{γ_0}. For this we have to restrict our choice for g_v a bit for vertices on the bottom and top of $\mathrm{I}\times\mathrm{I}$.
For γ_0 we can view vertices as elements in I, since they all have second coordinate 0 anyway. Furthermore there are two types of vertices. The ones that come from the factorization $[f_1]\cdots[f_m]$ (we added those in **Change 1** above), called **special**, and the rest.
Any non-special vertex v of γ_0 lies in a smallest interval $[v_l, v_r]$ between two special vertices. This is true since 0 and 1 are always special. For such an interval we know that $(f_1\cdots f_m)([v_l, v_r]) \subset \mathrm{A}_\alpha$ for some α and our choice of multiplication order. Since v is a vertex on the bottom of $\mathrm{I}\times\mathrm{I}$ it is contained in at most two rectangles R_i and R_j. So we can assume that the path g_v from before is in $\mathrm{A}_{\alpha(i)} \cap \mathrm{A}_{\alpha(j)} \cap \mathrm{A}_\alpha$. Then the part

$$[g_{v_l}\cdot f_{e_{l+1}}\cdot \overline{g}_{v_{l+1}}]\cdots[g_{v_{r-1}}\cdot f_{e_r}\cdot \overline{g}_{v_r}],$$

of the factorization for f_{γ_0} can be seen as having all factors in $\pi_1(\mathrm{A}_\alpha)$, by using modifications of type (3), and so by modifications of type (2) this product is equivalent to

$$[g_{v_l}\cdot f_{e_{l+1}}\cdots f_{e_r}\cdot \overline{g}_{v_r}].$$

Recall, $g_{v'}$ was a path from x_0 to $F(v')$ for a vertex v' (or $(f_1\cdots f_m)(v')$ when viewed as an element of the interval). But if v' is special then $(f_1\cdots f_m)(v') = x_0$. So $g_{v'}$ for a special vertex v' was chosen to be the constant path before. Thus the above expression simplifies to

$$[f_{e_{l+1}}\cdots f_{e_r}],$$

which is the homotopy class of $f_1\cdots f_m$ restricted to the interval $[v_l, v_r]$ and hence equal to the factor in $[f_1]\cdots[f_m]$ corresponding to the interval $[v_l, v_r]$.
Fix now all these choices of g_v, including the more restricted choices for the vertices on the bottom or top of $\mathrm{I}\times\mathrm{I}$. Then we obtain that $[f_1]\cdots[f_m]$ is equivalent to the factorization of f_{γ_0} (that was the last step), which in turn is inductively equivalent to the factorization of $f_{\gamma_{pq}}$ (that was the step before the last one), which is then equivalent to the factorization $[f_1']\cdots[f_n']$ (this is the same argument as the last step). In total we thus get that

$$[f_1]\cdots[f_m] \sim [f_1']\cdots[f_n'],$$

which in turn implies that the kernel of φ is exactly N. So part (b) is proven. $\square$