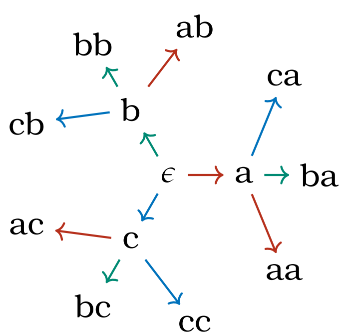
Please Distribute

An invitation to tensors

2024-03-24

James B. Wilson



Department of
Mathematics
Colorado State
University

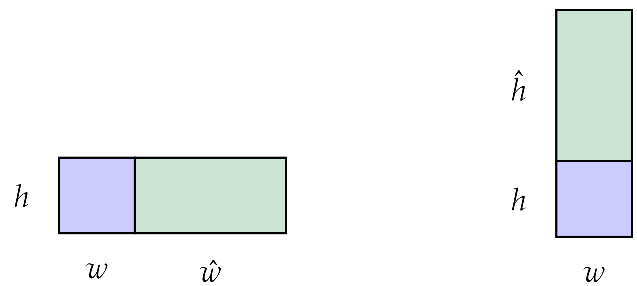
James.Wilson@ColoState.Edu

James.Wilson@ColoState.Edu

Contents

The distributive property 1

The distributive law will be the star of the tensor show. Distributive laws appear throughout our world of measurement and this offers some gentle points of entry to the study. Start with area.



So area as function $A(w,h)$ of width and height distributes over those parameters.

$$A(w+\hat{w},h) = A(w,h) + A(\hat{w},h) \qquad A(w,h+\hat{h}) = A(w,h) + A(w,\hat{h})$$

Calculus says that all areas are defined by distributed sums of subrectangles, or by the steady approximation by such rectangles. Thus area's distributive property water-marks calculus in numerous ways including the following fundamental identity.

$$\int_M (f+g)\mathbf{d}x = \int_M f\,\mathbf{d}x + \int_M g\,\mathbf{d}x$$

"Real" life

Look elsewhere in your life. When you buy x items at cost $\mathrm{Cost}(x)$ and then buy $\acute{x}$ more the total cost is:

$$\mathrm{Cost}(x+\acute{x}) = \mathrm{Cost}(x) + \mathrm{Cost}(\acute{x}).$$

Add the cost of tax $\mathrm{Tax}(x)$ and the total changes to

$$(\mathrm{Cost}+\mathrm{Tax})(x) = \mathrm{Cost}(x) + \mathrm{Tax}(x).$$

It is not just cost of one item. Buy items $x_1,\ldots,x_n$ at costs $C_i(x_i)$ each and set $\mathrm{Cost}(x_*) = \mathrm{Cost}(x_1,\ldots,x_n) = C_1(x_1)+\cdots+C_n(x_n)$ Still we find $\mathrm{Cost}(x_*+\acute{x}_*) = \mathrm{C}(x_*) + \mathrm{Cost}(\acute{x}_*)$. Cost distributes.

Here is another life situation. The work of e employee for h hours produces $W(e,h)$ widgets so at peak productivity adding $\hat{e}$ workers or $\hat{h}$ extra hours we get

$$W(e+\hat{e},h) = W(e,h) + W(\hat{e},h) \qquad W(e,h+\hat{h}) = W(e,h) + W(e,\hat{h}).$$

Work distributes, both over workers and over time.

Need to print the names of both faculty and graduate students. You could distribute that work as well, as seen in Listing 1.0.1, the command `Print` distributes over the operator of concatenation `cat` with the result of simply appending the two individual printing commands. We could say

```
Print(as cat bs)== Print(as)append Print(bs)
```

Listing 1.0.1 Many operations in computation are distributive.

In [1]
```
fac = [ "Hamilton", "Levi-Chivita" ]
students = [ TBD use math geoneolgoy to find some
    students of these]
Print(faculty cat students)
```

Out [1]
```
Hamilton
Levi-Chivita
....TBD
```

In [2]
```
Print(faculty)
Print(students)
```

Out [2]
```
Hamilton
Levi-Chivita
....TBD
```

Well all this is at least true for some range of values. All scientific claims of connection to reality are true only for sensible regions of values.

The distributive appears so often because we are willing to be flexible about what we want from addition and multiplication. We just use those two names to remember “multiplication distributes over addition”—whatever those two terms mean.

Notations

By custom, area $A(w,h)$ of rectangles is denoted $w \cdot h$ and called multiplication. In that notation we have the familiar form.

$$(w+\hat{w})\cdot h = w\cdot h + \hat{w}\cdot h \qquad w\cdot(h+\hat{h}) = w\cdot h + w\cdot\hat{h}.$$

This is the first of dozens of notations invented over the centuries to communicate functions that distribute their work amongst sums of sub-applications of the function. Most honor one of these three modes:

- functions on the outside applied to sums of inputs

  $f(u+\acute{u}) = f(u)+f(\acute{u})$, `Print(as cat bs)`, Cost$(x,y,z)\ldots;$

- functions in the middle like these operators

  $(u+\acute{u})*v = u*v+\acute{u}*v, \qquad (u+\acute{u})\cdot v = u\cdot v+\acute{u}\cdot v, \qquad u\otimes v, \qquad uv,\ldots;$

- functions on the outside like these

  $\langle u+\acute{u}|v\rangle = \langle u|v\rangle+\langle\acute{u}|v\rangle, \qquad [u+\acute{u},v] = [u,v]+[\acute{u},v], \qquad /u,v/,\ldots\ .$

Axes and dependent-functions

Just, as area distributes, so does volume $V(\ell,w,h)$ only with one more axis.

[TBD: a tikz picture of three volumes showing three axes of distributive volume]

As formulas this could be written:

$$\begin{aligned}
V(\ell+\hat{\ell},w,h) &= V(\ell,w,h)+V(\hat{\ell},w,h)\\
V(\ell,w+\hat{w},h) &= V(\ell,w,h)+V(\ell,\hat{w},h)\\
V(\ell,w,h+\hat{h}) &= V(\ell,w,h)+V(\ell,w,\hat{h})
\end{aligned}$$

With volumes we obviously must give up on the chiral language of “distributes on the left” “distributes on the right”. Venturing into hypervolumes will be worse and benefits from some notation.

First we need to name every coordinate and while we could number coordinates a for more flexible idea is the name them after what they mean. For example the rows and columns of matrix are typically used in that order and rows are interpreted as moving top-to-bottom.. Meanwhile x and y flip this convention with the left-to-right going first and bottom-to-top orientation for vertical. So a function M_{rc} in row-column form is differently interpreted from a function f_{xy} in x-axis, y-axis form. So naming the axes rather than prescribing fixed meaning is the most common.

Dependent-functions are elements of $\prod_{a\in A} U_a$.

Now that the axes A are named, when we want to changing value x_a on an axis $a \in A$ we do so merely by providing an annotation, usually a subscript. For example $A = \{\text{row}, \text{col}\}$ means we could consider $u_A = (u_{\text{row}}, u_{\text{col}})$. What is going on here is that we are not actually describing an ordered pair but instead a function whose output type depends on the input or what is more commonly known as a *dependent-function*. Order-pairs can be seen as special cases of dependent -functions, they depend on $A = \{\text{left}, \text{right}\}$ or sometimes $\{1, 2\}$, but the the convention for writing ordered pairs (x, y) is that the position of the value is fixed and from that one recovers the implicit input, e.g. x is in the left so implicitly left $\mapsto x$.

There are two notations used somewhat interhcangeably for dependent-functions

$$u_A : a \mapsto u_a \qquad\qquad u_A = (u_a \mid a \in A) \tag{1.1}$$

When it become tedious we write $u = u_A$. If we partition $A = B \sqcup C$ we can accordingly restrict the dependent-function u to the smaller domains creating two dependent-functions u_B and u_c with the same overall information, which we write

$$u_A = (u_B, u_C)$$

An especially important case is for $a \in A$ where we let $\bar{a} = A - \{a\}$ and write

$$u_A = (u_a, u_{\bar{a}})$$

Back to volume we could now have stated $A = \{\text{l}, \text{w}, \text{h}\}$ as the axes. Then volume is a function $\text{Vol}(u_A)$ where for each $a \in A$,

$u_a \geq 0$. Thus the distributive property in all three variables can be condensed to a single formula quantified over every axis.

$$(\forall a \in A) \qquad \mathrm{Vol}(u_a + \acute{u}_a, u_{\bar{a}}) = \mathrm{Vol}(u_a, u_{\bar{a}}) + \mathrm{Vol}(\acute{u}_a, u_{\bar{a}})$$

Notice if we change $A = \{\mathrm{w}, \mathrm{h}\}$ we get the distributive law for area, and if we change to $A = \{x, y, z\}$ we get volume but no in reference to xyz-coordinates instead of the more ambiguous length-width-height coordinates. Hypervolumes are then the same rule but with more axes such as $A = \{x, y, z, t\}$ or axes $A = \{x_1, \ldots, x_\ell\}$.

In these notes $u_A = (u_a \mid a \in A)$ *denotes a dependent-function not an ordered tuple.*

1.1 Distributive algebra

What does distribution require?

$$u * (v + \acute{v}) = u * v + u * \acute{v} \qquad (u + \acute{u}) * v = u * v + \acute{u} * v. \tag{1.2}$$

We definitely need additions, but we should not jump to assume that u, v, and w are of the same type of data. Just look at matrix multiplication (we use $\mathbb{R}^{a\times b}$ to denote $(a \times b)$-matrices of real numbers)

$$* : \mathbb{R}^{a\times b} \times \mathbb{R}^{b\times c} \to \mathbb{R}^{a\times c} \qquad (u * v)_{ij} = \sum_k u_{ik} v_{kj}.$$

So we except this is a study of *heterogeneous* algebra, so we wont be captivated by homomorphism but rather what we will call *hetero*morphisms. So we could think of three types of data U, V and W each with a $+$ each combined by a function $* : U \times V \to W$ that satisfies the distributive law. It turns out we often need a little more.

Proposition 1.1. If $+$ is a binary operation on U then we may add a term 0 disjoint form U, $U_0 := U \sqcup \{0\}$ with addition

$\dot{+}$	0	b
0	0	b
a	a	$a+b$

So every additive structure may be assumed to have a 0 where $x+0=x=0+x$.

Definition 1.2. Let U, V and W have additions $+$ (a binary operator) and a 0. A *bimap* (biadditive map) is a function $*: U \times V \rightarrowtail W$ where

$$
\begin{aligned}
(u+\acute{u})*v &= u*v+\acute{u}*v\\
u*(v+\acute{v}) &= u*v+u*\acute{v}\\
0*v = 0 &= u*v.
\end{aligned}
$$

As we go along we will prefer to use U, V and W in just this way so that we can get up to speed on examples as quickly as possible.

Properties of addition

It is tempting now to start assuming that U, V, and W are something family—vector spaces, modules, or at least abelian groups. However this would rob the distributive law of its power and leave us to think addition and its common attributes are the reason tensors work. But the distributive law is already claiming a strong interaction of two operations so maybe it should be explored on its own a little while longer to appreciate what it already says about the individual operations. More examples will demonstrate the value of a general point of view.

We will use a number of spaces

$$\begin{aligned}
\mathbb{R}^d &:= \{u : [d] \to \mathbb{R}\},\\
R^{m\times n} &:= \{M : [m]\times[n] \to \mathbb{R}\}\\
R^{\ell\times m\times n} &:= \{\Gamma : [\ell]\times[m]\times[n] \to \mathbb{R}\}\\
&\vdots
\end{aligned}$$

Define the following operations.

$$\mathbb{R}^m \oplus \mathbb{R}^n := \mathbb{R}^{m+n}$$
$$\begin{bmatrix}\mathbb{R}^{a\times n}\\ \mathbb{R}^{b\times n}\end{bmatrix} := \mathbb{R}^{(a+b)\times n} \qquad \begin{bmatrix}\mathbb{R}^{m\times c} & \mathbb{R}^{m\times d}\end{bmatrix} := \mathbb{R}^{m\times(c+d)}$$
$$\vdots$$

Now we add a multiplication.

$$\mathbb{R}^m \otimes \mathbb{R}^n := \mathbb{R}^{m\times n}$$

Example 1.3. The distributive law with vector space operators.

$$(\mathbb{R}^a \oplus \mathbb{R}^b)\otimes \mathbb{R}^n = \begin{bmatrix}\mathbb{R}^a\otimes\mathbb{R}^n\\ \mathbb{R}^b\otimes\mathbb{R}^n\end{bmatrix}$$
$$\mathbb{R}^m \oplus (\mathbb{R}^c \oplus \mathbb{R}^d) = \begin{bmatrix}\mathbb{R}^m\otimes\mathbb{R}^c & \mathbb{R}^m\otimes\mathbb{R}^d\end{bmatrix}$$

Lets combine the left and right distributive laws.

$$\begin{array}{ccccc}
 & & (u+\acute{u})*(v+\acute{v}) & & \\
 & // & & \backslash\backslash & \\
(u+\acute{u})*v+(u+\acute{u})*\acute{v} & & & & u*(v+\acute{v})+\acute{u}*(v+\acute{v})\\
\| & & & & \|\\
(u*v+\acute{u}*v)+(u*\acute{v}+\acute{u}*\acute{v}) & & & & (u*v+u*\acute{v})+(\acute{u}*v+\acute{u}*\acute{v})
\end{array}$$

Thus the values $a,b,c,d,\ldots$ in the image of a distributive product $*: U\times V \rightarrowtail W$ must satisfy the following identity.

$$(a+b)+(c+d) = (a+c)+(b+d) \qquad \text{(Distributable)}$$

Proposition 1.4. If $+$ is both associative and commutative then it is distributable. Furthermore, such an addition can be extended to have at 0.

Conversely, if $+$ is distributable and has a 0 then $+$ is associative and commutative.

Proof. TBD □

1.2 Commonoids are required

$a \equiv \acute{a}\ (R)$ is a congruence relation of an algebra.

$\mathrm{Fr}_\Omega\langle S\rangle$ is the subalgebra of evaluate all Ω-formulas $\Phi(X)$ at values in S. We say S *generates* this subalgebra.

Now we get to tame the algebra.

Definition 1.5. Given a distributive product $*: U \times V \rightarrowtail W$ let

$$\begin{aligned}(u \equiv \acute{u} \quad (V^\top)) &\Leftrightarrow (\forall v \in V)(u * v = \acute{u} * v)\\ (v \equiv \acute{v} \quad (U^\perp)) &\Leftrightarrow (\forall u \in U)(u * v = u * \acute{v})\\ U * V &:= \mathbf{Fr}_{+,0}\langle u * v \mid u \in U, v \in V\rangle\end{aligned}$$

When necessary we denoted equivalence classes $u/V^\perp$ and etc.

Proposition 1.6. Each of the following induced products are well-defined and distributive.

- $*: U/V^\top \times V \rightarrowtail W$
- $*: U \times V/U^\perp \rightarrowtail W$
- $*: U \times V \rightarrowtail W^+$
- $*: U/V^\top \times V/U^\perp \rightarrowtail U * V$

Proof. TBD by a Wilsonite □

Definition 1.7. An additive algebra that is both associative and commutative is called a *commonoid*.

Proposition 1.8. Given a distributive product $*: U \times V \to W$ then $U/V^\perp$, $V/U^\top$ and $W^+ = U * V$ are commonoids.

Proof. TBD by a Wilsonite □

In commutative diagram form (i.e. flow charts where any two paths between the same points agree) we have the following.

$$\begin{tikzcd} U \arrow[d, two heads] & \times & V \arrow[d, two heads] \arrow[r, tail, "*"] & W & \text{(Additive algebras)} \\ U/V^{\top} & \times & V/U^{\perp} \arrow[r, tail, "\bullet"] & U*V \arrow[u, hook, "h"] & \text{(Commonoids)} \end{tikzcd}$$

We have surjections and inclusion so something is lost, but none of it is information! Notice $u \equiv \acute{u} \;\; (V^{\top})$ precisely when for *every* $v \in V$, $u*v = \acute{u}*v$. So factoring this in does not influence those values. Likewise with moding out by $U^{\perp}$. Finally, $U*V$ is generated by the image of the function $*$ so anything in $W - U*V$ is something that the function ignores.

This says that every every bimap induces one that is defined commonoids and no information is lost.

[TBD: describe Tucker decomposition]

Grothendieck group

Theorem 1.9. If M is a commonoid then

$$(m,n) \equiv (\acute{m},\acute{n}) \quad (Grth) \Leftrightarrow (\exists k \in M)(m+\acute{n}+k = \acute{m}+n+k)$$

is a congruence on $M \oplus M$. Furthermore,

$$(m,n)+(n,m) \equiv (0,0) \quad (Grth)$$

So $\mathrm{Gr}(M) := M \oplus M/Grth$ is an abelian group where

$$-(m,n) := (n,m)$$

This is called the Grothendieck group of M. The interpretation is that $(m,n)/R$ is “m-n”. When $m=0$ it is customary to write $-n$.

The homomorphism $M \to \mathrm{Gr}(M)$ given by $m \mapsto (m,0)$ is injective if, and only if, $m+n = m+\acute{n}$ implies $m = m$, i.e. M has cancellation.

Proposition 1.10. If $u \in U$ with $-u$ such that $u + (-u) = 0$ then

$$u * v + (-u) * v = 0$$

Thus if U has negatives then we can promote these negatives to $U * V$. In particular $-(u*v) := (-u)*v$ defines a negative in W of $u*v$. We can do likewise in the variable V.

Now imagine the distributive law was present along side an existing subtraction. Then the following would be a consequence.

$$\begin{aligned}(a-b)*(c-d) &= (a*(c-d)) - (b*(c-d)) = (a*c - a*d) - (b*c - b*d)\\ &= (a*c + b*d) - (a*d + b*c)\end{aligned}$$

This should explain the following extension.

Corollary 1.11. If $* : U \times V \rightarrowtail W$ is a distributive product of commonoids then there is an induced distributive product

$$\begin{gathered} * : \mathrm{Gr}(U) \times \mathrm{Gr}(V) \rightarrowtail \mathrm{Gr}(W)\\ (u - \acute{u}) * (v - \acute{v}) := (u * v + \acute{u} * \acute{v}) - (u * \acute{v} + \acute{u} * v). \end{gathered}$$

In particular $(-u) * v = -(u * v) = u * (-v)$.

Let us stand back to see what we have built.

$$\begin{tikzcd}
U \arrow[d, two heads] & \times & V \arrow[d, two heads] \arrow[r, tail, "*"] & W & \text{(Additive algebras)} \\
U/V^{\top} \arrow[d] & \times & V/U^{\perp} \arrow[d] \arrow[r, tail, "\bullet"] & U * V \arrow[u, hook, "h"] \arrow[d] & \text{(Commonoids)} \\
\mathrm{Gr}(U/V^{\top}) & \times & \mathrm{Gr}(V/U^{\perp}) \arrow[r, tail, "\bullet"] & \mathrm{Gr}(U * V) & \text{(Abelian Groups)}
\end{tikzcd}$$

It is not a mistake that in the last component the arrows are not aligned. This is actually a symptom of an unresolved problem in the study of tensors. They do not lie in a single category.

1.3 The tensor lifting property

1.4 The tensor product

Algebra

Formulas and subalgebras

We are in a territory where we are assuming very little about our addition so that we leave open the possiblity of applying the ideas of tensors broadly. This already included examples where the data we add include things like vector spaces, and there is no “set of all vector spaces”. In fact sets themselves are a bit of mismatch with algebra. This causes no trouble. Algebra predates sets (think of Boolean, Heyting and de Morgan algebra) and outlasts sets (think of categories). If this makes you uncomfortable you should note that programming is in the same boat. It describes logic outside of sets all the time. So what is algebra without sets? Well what is math without sets! It takes a step back and becomes all about language. In otherwords it is time to go back to grammar school.

We want additive structures so we use + and we want that be $x+y$, because at least for math styles like $+x\ y$ or `add(x,y)` are not our preference. To feel sophisticated about this we could report that we have the following grammar:

```
<A>::= (<A>+<A>) | 0
```

where | indicates “or” and so we can add or use `0`. If we do not want `0` in our addition we just remove it; likewise, if wish to add something like negatives we can add this as well as in the following.

```
<An>::= (<AN>+<AN>) | 0 | -<AN>
```

We will want some formulas. So we add variables, for examples to add $X=\{x,y,z\}$ as variables we use

```
   <X> ::= x | y
<A<X>> ::= <X> | 0 |(<A<X>>+<A<X>>)
```

Strings $\Phi := \Phi(X)$ that prase in this grammar are called *additive formulas* and annotated $\Phi : \mathrm{Fr}_A\langle X\rangle$. For example $((x+0)+y) : \mathrm{Fr}_A\langle x, y\rangle$, but not $+xy$ because that does not match the pattern in the grammar.

Given an additive structure W and assignment of variables $w_X : W^X$ we can substitute into a formula $\Phi(X)$ the values w_X formula a list of operations to carry out in W called *evaluation* and denoted $\Phi(w_X)$. For example $\Phi(x, y) = ((x+0)+y)$ and $W = \mathbb{N}$ then $\Phi(2, 3) = ((2+0)+3) = 5$.

Now given an additive algebra W and a function $S : I \to W$, then the evaluations $\Phi(s_*)$ are said to have type

$$\mathrm{Fr}_{+,0}\langle S\rangle$$

We say that $\mathrm{Fr}_{+,0}\langle S\rangle$ is the subalgebra generated by S.

Congruence, quotients, & homomorphisms

Algebra is the study of equations with variables, e.g. $x^2+1 = 0$. The most general form of algebra should allow every symbol to vary. So we can replace + one day with addition of decimal numbers, the next day + is addition of complex. Later we are adding matrices or polynomails etc. The + is changing. But so too is the =. Equality of decimals is not equality of matrices. The theory of changing equality is Noether's Isomorphism theorem. To get there let us recall an even more basic resonance.

Theorem .12 (Resonance Theorem)**.** In a calculus of constructions

$$\begin{tikzcd}
\text{Equivalence} \arrow[rr, bend left] & & \text{Partition} \arrow[dl, Rightarrow, bend left] \\
& \text{Functions} \arrow[ul, Rightarrow, bend left] &
\end{tikzcd}$$

Now we add algebra to the data and select those equivalence relations that respect the algebra. Equivalence relations that respect algebra are called *congruences*. By convention a congruence relation R is denoted as

$$a \equiv \acute{a} \quad (R)$$

where R is the name of the congruence.

Example .13. In a $\{+, 0, -\}$-algebra,

$$\begin{array}{rrcll} & a & \equiv & \acute{a} & (R) \\ + & b & \equiv & \acute{b} & (R) \\ \hline & a+b & \equiv & \acute{a}+\acute{b} & (R) \end{array} \qquad \begin{array}{cccc} \\ \hline 0 & \equiv & 0 & (R) \end{array} \qquad \begin{array}{rrcll} - & a & \equiv & \acute{a} & (R) \\ \hline & -a & \equiv & -\acute{a} & (R) \end{array}$$

Of course operators that depend on no parameters—0-valent operators—are trivially respected by all equivalence relations, but we include these in the list to make it clear that every operator is being considered.

A partition $\{[a] \mid a \in A\}$ of an algebra that respects the operations is called a *quotient*. For instance for a $\{+, 0, -\}$-algebra the partition would mean the following are well-defined.

$$[a] + [b] = [a + b] \qquad [0] = [0] \qquad -[a] = [-a].$$

By convention, quotients $\mathcal{P}$ are denoted $A/\mathcal{P}$ and the classes $[a]$, which we denote by $a/\mathcal{P}$, are called *cosets* (even when they are not sets).

Finally a function $\varphi : A \to B$ between algebras that respects operations of (homogeneous) algebra A will called a *homomorphism* and to complete the example here is what respect means for $\{+, 0, -\}$-algebras.

$$\varphi(a + \acute{a}) = \varphi(a) + \varphi(\acute{a}) \qquad \varphi(0) = 0 \qquad \varphi(-a) = -\varphi(a).$$

Now we update the resonance to match.

Theorem .14 (Noetherian Resonance)**.** In a Martin-Löf typed algebra

$$\begin{tikzcd}
\text{Congruence} \arrow[rr, Rightarrow, bend left] & & \text{Quotient} \arrow[dl, Rightarrow, bend left] \\
& \text{Homomorphism} \arrow[ul, Rightarrow, bend left] &
\end{tikzcd}$$

Proof. From congruences to quotients use:

$$a \equiv \acute{a} \;\; (R) \longrightarrow \begin{array}{l} A/R = \{a/R \mid a \in A\} \\ a/R := \{\acute{a} \mid a \equiv \acute{a} \;\; (R)\} \end{array}$$

From quotients to homomorphisms use:

$$A/\mathcal{P} = \{a/\mathcal{P} \mid a \in A\} \longrightarrow (a \mapsto a/\mathcal{P}) : A \to A/\mathcal{P}$$

From homomorphism to congruence use kernels.

$$\varphi : A \to B \longrightarrow a \equiv \acute{a} \;\; (\ker \varphi) \Leftrightarrow \varphi(a) = \varphi(\acute{a})$$

□