**Section 9: Equivalence Relations; Cosets**

**Def:** A *(binary) relation* on a set $S$ is a subset $\mathcal{R}$ of $S \times S$. If $(a, b) \in \mathcal{R}$, we write $a\mathcal{R}b$ and say "$a$ is $\mathcal{R}$-related to $b$".

Like the definition of a function as a set of ordered pairs, this definition conveys very little sense of what a relation is. It is really intended to serve as a "verb" in a sentence with subject and direct object elements of $S$; if $(a, b) \in \mathcal{R}$, then the sentence $a\mathcal{R}b$ is true, while if $(a, b) \notin \mathcal{R}$, then the sentence $a\mathcal{R}b$ is false. Our text hurries on to the particular kind of relation that we are presently interested in, namely an equivalence relation, but there are other kinds of relations. One of which I am fond is a "partial order relation", like "is a subset of" among subsets of a given set. Just to give an example of a relation, let's take the family $\mathcal{P}(A)$ of subsets of the set $A = \{b, c\}$: $\mathcal{P}(A) = \{\emptyset, \{b\}, \{c\}, A\}$, and the partial order relation "is a subset of" on $\mathcal{P}(A)$ is

$$\subseteq = \{\ (\emptyset, \emptyset),\ (\emptyset, \{b\}),\ (\emptyset, \{c\}),\ (\emptyset, A),\ (\{b\}, \{b\}),\ (\{b\}, A),\ (\{c\}, \{c\}),\ (\{c\}, A),\ (A, A)\ \}$$

In other words, the sentences $\emptyset \subseteq \{b\}$ and $\{c\} \subseteq A$ are true, but $\{b\} \subseteq \{c\}$ and $A \subseteq \{b\}$ are false. A completely artificial example would be $S = \{1, 2, 3\}$ and $\mathcal{R} = \{(1, 1), (1, 3), (2, 2), (2, 1)\}$. Here the statements $1\mathcal{R}3$ and $2\mathcal{R}1$ are true, but $2\mathcal{R}3$ is not true.

But let's get back to the focus of the section: equivalence relations, i.e., relations which say that two things in a set have "equal value" from some point of view — not that they are really identical (although one property will say that if they are identical, then they are equivalent), but that they can be grouped together and treated as the same under certain circumstances. Here are the official rules:

**Def:** A relation $\mathcal{R}$ on a set $S$ is an *equivalence relation* if it satisfies the following conditions:

(R) it is *reflexive,* i.e., for all $s$ in $S$, $s\mathcal{R}s$ (in other words, every element of $S$ is related to itself — or, in still other words, the "diagonal" $\{(s, s) : s \in S\}$ is a subset of $\mathcal{R}$);

(S) it is *symmetric,* i.e., if $s\mathcal{R}t$ then also $t\mathcal{R}s$; and

(T) it is *transitive,* i.e., if $s\mathcal{R}t$ and $t\mathcal{R}u$, then also $s\mathcal{R}u$.

**Ex:** In $\mathbb{Z}$, "has the same remainder on long division by 5" is an equivalence relation. "Has the same sign as" is an equivalence relation only if we say that the sign of 0 is something (because 0 has to be related to itself) but it can't be both positive and negative, because then transitivity would say 1 and $-1$ have the same sign.

**Ex:** In geometry, congruence of triangles is an equivalence relation. So is similarity of triangles (same shape, different sizes). "Has the same area as" would work, too.

There are lots of other examples, but before proceeding, let's look at the following almost equivalent ways of thinking about equivalence relations:

**Def and Prop:** (1) A *partition* of a set $S$ is a family $\mathcal{P}$ of nonempty subsets of $S$ with the property that every element of $S$ is in exactly one of the sets in $\mathcal{P}$; i.e., for all $s$ in $S$, there is an $A$ in $\mathcal{P}$ for which $s \in A$, but if $A, B \in \mathcal{P}$ and $A \neq B$, then $A \cap B = \emptyset$. If $\mathcal{P}$ is a partition of $S$, then "is in the

same element of $\mathcal{P}$" is an equivalence relation on $S$.

(1′) Conversely, suppose $\mathcal{R}$ is an equivalence relation on $S$. For each $s$ in $S$, we set

$$[s] = \{t \in S : t\mathcal{R}s\} \ .$$

(If there are more than one equivalence relation under consideration, it may be necessary to add the symbol $\mathcal{R}$: $[s] = [s]_{\mathcal{R}}$. The text denotes $[s]$ by overbar-$s$, but I wanted something easier to read and something that looked more like a set.) It is called the *$\mathcal{R}$-equivalence class of $s$.* Then the set of $\mathcal{R}$-equivalence classes is a partition of $S$.

(2) If $f : S \to T$ be a function on $S$, then "has the same image under $f$" is an equivalence relation on $S$.

(2′) Conversely, if $\mathcal{R}$ is an equivalence relation on $S$, then the assignment $s \mapsto [s]$ is a function from $S$ to the set of $\mathcal{R}$-equivalence classes.

*Pf of whatever is not a definition here:* (1) Obviously every element of $S$ is in the same element of $\mathcal{P}$ as itself, so this relation is reflexive. And if $s, t$ are in the same element of $\mathcal{P}$, then $t, s$ are in the same element of $\mathcal{P}$, so it is symmetric. Transitivity follows from the fact that the elements of $\mathcal{P}$ don't overlap: If $s, t$ are in the same element of $\mathcal{P}$, and $t, u$ are in the same element of $\mathcal{P}$, then $s, u$ are in the same element of $\mathcal{P}$.

(1′) None of the $\mathcal{R}$-equivalence classes are empty because of reflexivity: $s\mathcal{R}s$, so $s \in [s]$. And for the same reason, every element of $S$ is in at least one of $\mathcal{R}$-equivalence classes, namely its own. Suppose two of the equivalence classes have non-empty intersection, say $[s] \cap [t] \neq \emptyset$; then we must show that $[s] = [t]$: Take $u$ in that intersection. Then by definition of equivalence class, we have $u\mathcal{R}s$ and $u\mathcal{R}t$. By symmetry we have $s\mathcal{R}u$ and then by transitivity we have $s\mathcal{R}t$. Thus, for all $w$ in $[s]$, we have $w\mathcal{R}s$, hence $w\mathcal{R}t$, hence $w \in [t]$. Therefore, $[s] \subseteq [t]$. Reversing the roles of $s$ and $t$ in the last four sentences, we get $[t] \subseteq [s]$ also, so $[s] = [t]$.

This proof shows that a single equivalence class can be written in many different ways, represented by any element in it: $t \in [s]$ iff $[t] = [s]$.

(2) and (2′) These are easy.//

**Inspiring example:** Instead of thinking of $0, 1, 2, 3, 4$ as elements of $\mathbb{Z}$, we could think of them as representing the "remainder classes on division by 5":

$$[3] = \{\ldots, -7, -2, 3, 8, 13, \ldots\}$$

because $-7 = -2(5) + 3$, $-2 = -1(5) + 3$, $3 = 0(5) + 3$, $8 = 1(5) + 3$, $13 = 2(5) + 3$, etc. In this sense $\mathbb{Z}_5$ is the set of remainder classes on division by 5, a partition of $\mathbb{Z}$. And we can think of the function that sends each integer to its remainder on division by 5, a function from $\mathbb{Z}$ to $\mathbb{Z}_5$.

The rest of the content of this section is to generalize that last example: Given a subgroup $H$ of a group $G$, we will partition $G$ into pieces, each with the same cardinality as $H$. (We will always get a function that sends each group element to its set in the partition; but later we will see that we can't always make the partition into a group like $\mathbb{Z}_5$.) The partition will allow us to get some nice counting formulas in a finite group.

**Prop:** Let $H$ be a subgroup of a group $G$. Then the relation defined by $a\mathcal{R}b$ iff $ab^{-1} \in H$ is an equivalence relation on $G$, with equivalence classes $Ha = \{ha : h \in H\}$, the "right cosets of $H$ in $G$." Similarly, $a^{-1}b \in H$ is an equivalence relation on $G$ with equivalence classes the "left cosets" $aH$.

*Pf:* We leave it to the reader to prove the last sentence, by mimicry.

Equivalence relation: Reflexive because $aa^{-1} = e \in H$ for all $a$ in $G$. Symmetric because if $ab^{-1} \in H$, then $ba^{-1} = (ab^{-1})^{-1} \in H$. Transitive because if $ab^{-1} \in H$ and $bc^{-1} \in H$, then $ac^{-1} = (ab^{-1})(bc^{-1}) \in H$.

Equivalence class as specified: If $b \in [a]$, i.e., $b\mathcal{R}a$, i.e., $ba^{-1} \in H$, then $b = (ba^{-1})a \in Ha$. If $b \in Ha$, i.e., $b = ha$ where $h \in H$, then $ba^{-1} = h \in H$, so $b\mathcal{R}a$, so $b \in [a]$. Therefore, $[a] = Ha$.//

Note that $H$ is one of its own cosets (left or right); it contains the identity, so naturally enough $He = H$. The other cosets are not subgroups — they don't contain the identity, for example. It may be useful to think of the cosets of $H$ as "translates" of $H$ within $G$, just as all the planes in $\mathbb{R}^3$ that are parallel to a given plane $P$ through the origin; $P$ is a vector subspace of $\mathbb{R}^3$, and the other planes parallel to $P$ are not subspaces.

It follows that the right cosets of $H$ in $G$ form a partition of $G$ (and similarly for the left cosets). Moreover, each right coset has the same cardinality as $H$ itself, because $H \to Ha : h \mapsto ha$ is one-to-one and onto. Let's look at some examples of cosets partitioning some groups:

**Ex:** The cosets of $5\mathbb{Z}$ in $\mathbb{Z}$ (left or right — they are the same because $\mathbb{Z}$ is abelian) are

$$\begin{array}{ll} 5\mathbb{Z} & : \ \ldots, -5, 0, 5, 10, 15, \ldots \\ 5\mathbb{Z}+1 & : \ \ldots, -4, 1, 6, 11, 16, \ldots \\ 5\mathbb{Z}+2 & : \ \ldots, -3, 2, 7, 12, 17, \ldots \\ 5\mathbb{Z}+3 & : \ \ldots, -2, 3, 8, 13, 18, \ldots \\ 5\mathbb{Z}+4 & : \ \ldots, -1, 4, 9, 14, 19, \ldots \end{array}$$

**Ex:** The cosets of $\langle 3 \rangle$ in $\mathbb{Z}_{15}$ are

$$\begin{array}{ll} \langle 3 \rangle & : \ 0, 3, 6, 9, 12 \\ \langle 3 \rangle \oplus 1 & : \ 1, 4, 7, 10, 13 \\ \langle 3 \rangle \oplus 2 & : \ 2, 5, 8, 11, 14 \end{array}$$

**Ex:** The right cosets of $\langle f \rangle$ in $D_5$ are

$$\begin{array}{ll} \langle f \rangle & : \ e, f, f^2, f^3, f^4 \\ \langle f \rangle g & : \ g, fg, f^2g, f^3g, f^4g \end{array}$$

while the left cosets of $\langle f \rangle$ in $D_5$ are

$$\begin{array}{ll} \langle f \rangle & : \ e, f, f^2, f^3, f^4 \\ g\langle f \rangle & : \ g, gf = f^4g, gf^2 = f^3g, gf^3 = f^2g, gf^4 = fg \end{array}$$

Of course, these are the same sets: One coset (left or right) is $\langle f \rangle$ itself and the other coset is the other half of $D_5$. But the right cosets of $\langle g \rangle$ in $D_5$ are

$$\begin{array}{ll} \langle g \rangle & : \ e, g \\ \langle g \rangle f & : \ f, gf = f^4g \\ \langle g \rangle f^2 & : \ f^2, gf^2 = f^3g \\ \langle g \rangle f^3 & : \ f^3, gf^3 = f^2g \\ \langle g \rangle f^4 & : \ f^4, gf^4 = fg \end{array}$$

while the left cosets of $\langle g \rangle$ in $D_5$ are

$$\begin{array}{rl} \langle g \rangle : & e \ , \ g \\ f\langle g \rangle : & f \ , \ fg \\ f^2\langle g \rangle : & f^2 \ , \ f^2 g \\ f^3\langle g \rangle : & f^3 \ , \ f^3 g \\ f^4\langle g \rangle : & f^4 \ , \ f^4 g \end{array}$$

Here the right and left cosets are different: $f^4 g \in \langle g \rangle f$, but $f^4 g \notin f\langle g \rangle$.

**Building $\mathbb{R}$:** At this point I would like to take a side trip, to show how the idea of equivalence relation was used (admittedly, not explicitly, because it wasn't invented; but the basic ideas were in there somewhere) to build the real number system. Kronecker (1823-91) said, "God made the integers; all else is the work of man;" but I think that is misleading. My sense is that God made humankind in such a way that humans would be compelled to invent the integers for themselves. In nature one finds lots of examples of three things (three rocks, three sticks, three wildebeests), but the independent concept of "three" is an abstraction from these examples, and hence a product of the human mind.

So let's put ourselves in the mathematical position of a person in, say, the Bronze Age: we know about the natural numbers

$$\mathbb{N} = \{1, 2, 3, \dots\}$$

(in modern notation), and we know all we need to know about adding them and multiplying them: A pile of 3 rocks combined with a pile of 2 rocks gives the same result as combining piles of 2 and 3; so we know addition is commutative. If we make two rectangles of rocks, both with 7 rocks in each row, and we end up with the same number of rocks in both rectangles, we know that the rectangles must both have had the same number of rows; so we know cancellation is possible in multiplication. But now we start working with sticks, which we can break into pieces; and we'd like to have new concepts that tell us how much wood we have in broken pieces of sticks. In other words, we want to invent the positive rationals $\mathbb{Q}^+$.

We might just make up new symbols like $2/3$ — it wouldn't be hard to imagine using the fraction line to "chop" a stick into three pieces and then taking two of them. But then we have the problem of why does one symbol, $4/6$, mean just as much wood as, say, $6/9$? Somehow we want to say that these two symbols have equal "wood value", in terms of addition and multiplication of natural numbers. It would be an amazing insight, the discovery of a Bronze Age Newton, to realize that a $4 \times 9$ rectangle of rocks has the same number as a $6 \times 6$ rectangle; and that if we think of a pile of 36 rocks as representing a stick, we can "chop" it into 6 equal piles and take four of them, or into 9 equal piles and take 6 of them; and end up with the same number of rocks. But with this discovery, we have the basis for deciding when two of these symbols have the same "wood value": $a/b = c/d$ when $ad = cb$. (Again, modern notation generalizing the ancient idea.)

But will this version of "equal wood value" make sense? In the terminology of this section, because these symbols have numerator and denominator from $\mathbb{N}$, we are asking whether, on the set $\mathbb{N} \times \mathbb{N}$, is "$(a, b)\mathcal{R}(c, d)$ means $ad = cb$" defines an equivalence relation. (Maybe the letter $\mathcal{R}$ stands for "ratio.") If so, $a/b$ can represent the equivalence class of all pairs with this "value." So, is $\mathcal{R}$ an equivalence relation?

(R) Is it true that, for all $(a, b)$, we have $(a, b)\mathcal{R}(a, b)$? This last relation is just $ab = ab$, which is true. So this $\mathcal{R}$ is reflexive.

(S) If $(a, b)(c, d)$, must we have $(c, d)\mathcal{R}(a, b)$? The hypothesis means $ad = cb$, while the conclusion means $cb = ad$, so the hypothesis implies the conclusion: $\mathcal{R}$ is symmetric.

(T) If $(a,b)(c,d)$ and $(c,d)(e,f)$, must we have $(a,b)(e,f)$? The hypothesis means $ad = cb$ and $cf = ed$, while the conclusion means $ae = fb$. This is harder: We can multiply the first two equations to get $adcf = cbed$, we noted earlier that we have cancellation of multiplication, so we can cancel $c$ and $d$ on both sides and get the conclusion. (We also used commutativity of multiplication, but that is pretty clear from rectangles of rocks.) So $\mathcal{R}$ is transitive.

So we have succeeded in building all the ratios of natural numbers, as $\mathcal{R}$-equivalence classes; i.e., we have built $\mathbb{Q}^+$. But we want it to be more than just a set of symbols; we want to be able to add and multiply them, too, as if they really were numbers. It must have taken a couple of millenia to find the formulas for operating with these symbols, but we don't have to wait:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+cb}{bd} \qquad \text{and} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \ .$$

Now, however, we do have a problem, which arises any time there are different ways to write an element of the domain of a function: Are these operations "well-defined", in the sense that, if we write the operands in a different way, do we get the same result? After all, the formula is in terms of $a, b, c, d$. If we change $a/b$ to the equal fraction $a'/b'$ and/or similarly for $c/d$, will the new answer be the same as the old? If we were to try to define addition of fractions the "easy way",

$$\frac{a}{b} + \frac{c}{d} = \frac{a+c}{b+d} \ ,$$

then, even through $1/2 = 2/4$, we have

$$\frac{1}{2} + \frac{1}{3} = \frac{2}{5} \ , \quad \text{but} \quad \frac{2}{4} + \frac{1}{3} = \frac{3}{7} \neq \frac{2}{5} \ .$$

So let's check that the <u>real</u> addition formula about <u>is</u> well-defined: Suppose $a/b = a'/b'$ and $c/d = c'/d'$; we want to show that $(ad+cb)/bd = (a'd'+c'b')/b'd'$. Now the hypothesis means $ab' = a'b$ and $cd' = c'd$, and we need to show that $(ad+cb)b'd' = (a'd'+c'b')bd$. But this last equation is just

$$adb'd' + cbb'd' = a'd'bd + c'b'bd \ ,$$

and using the hypothesis (and commutativity of multiplication in $\mathbb{N}$ again), we can get that the left side of the displayed equation is equal to the right:

$$\begin{aligned} adb'd' + cbb'd' &= (ab')(dd') + (cd')(bb') \\ &= (a'b)(dd') + (c'd)(bb') \\ &= a'd'bd + c'b'bd \ . \end{aligned}$$

Checking that multiplication is also well-defined is a little easier.

So a few millenia later, positive rational numbers have become routine, and there is a vague sense of the number 0. But negative numbers are still "imaginary" things, and some questions have no answers: What, added to 5, gives 3? So how can we build $\mathbb{Q}$ from $\mathbb{Q}^+$? The answer is again, $\mathbb{Q}$ is the set of equivalence classes carved out of $\mathbb{Q}^+ \times \mathbb{Q}^+$ by an equivalence relation. This time the relation is $\mathcal{D}$, (maybe for "difference"): We want two ordered pairs $(a,b), (c,d)$ to be related if they represent the same difference, positive or negative: $a - b = c - d$, or equivalently, $a + d = c + b$. It is left to the reader to show that this is an equivalence relation on $\mathbb{Q}^+ \times \mathbb{Q}^+$. So now we can say

- the difference $a - b$ is the equivalence class containing $(a,b)$ — so now every difference makes sense, whether $a > b$ or $a < b$;

- 0 (which was sort of a shadowy nonentity so far) is the equivalence class containing all the $(a, a)$'s; and
- $-a$ is the equivalence class containing $(1, a+1)$, or equivalently $(b, a+b)$ for any $b$.

There is more to do, of course, like defining the operations on these new "numbers"; but now we have all of $\mathbb{Q}$.

The big jump, from rationals to reals, was actually taken before the jump from positives to negatives. Eudoxus explained how to construct the positive reals $\mathbb{R}^+$ from the positive rationals $\mathbb{Q}^+$, long enough ago that Euclid could include it in the *Elements* in about 300 BCE. And two millenia later, Dedekind copied the process to build $\mathbb{R}$ from $\mathbb{Q}$. The idea really fits better in Math 323, but here is a sketch: A "Cauchy sequence" of rational numbers is an infinite list of numbers $(a_n) = (a_1, a_2, a_3, \dots)$ in which the $a_n$'s are getting close to each other as $n$ gets larger — trying to converge to a real number, which may or may not be there. Here's an example:

$$1 \quad 1.4 \quad 1.41 \qquad 1.414 \qquad 1.4142 \quad \dots$$

all rational, but "trying to get close to" $\sqrt{2}$. On the set of all rational Cauchy sequences, we define the relation $\mathcal{L}$ (maybe for "limit") by $(a_n)\mathcal{L}(b_n)$ if the $a_n$'s are getting closer to the $b_n$'s as $n$ get's larger — i.e., if they have the same limit. The $\mathcal{L}$-equivalence classes of rational Cauchy sequences are the real numbers.

**Interlude: "Is conjugate to" in two contexts**

Recall that, for elements $x, y$ of a group $G$, we say that $y$ is "conjugate" to $x$ (in $G$) if there is an element $g$ of $G$ for which $gxg^{-1} = y$. Notice that "is conjugate to" is an equivalence relation on $G$. (Proof: Every element is the conjugate of itself by the identity. If $y$ is conjugate to $x$ by $g$, then $x$ is conjugate to $y$ by $g^{-1}$. If $y$ is conjugate to $x$ by $g$ and $z$ is conjugate to $y$ by $h$, then $z$ is conjugate to $x$ by $hg$.) This may seem like a rather formal and abstract way for group elements to be related. We want to note here how, in two specific groups, "is conjugate to" has a natural meaning.

First, in any of the groups $\mathcal{S}_n$, for any $r$-cycle $(t_1, t_2, \dots, t_r)$ (where the $t_i$'s are, of course, elements of $\{1, 2, \dots, n\}$), and any element $\varphi$ of $\mathcal{S}_n$, we have, for each $x \in \{1, 2, \dots, n\}$:

$$\begin{aligned}
(\varphi \circ ((t_1, t_2, \dots, t_r) \circ \varphi^{-1})(x) &= \varphi((t_1, t_2, \dots, t_r)(\varphi^{-1}(x)) \\
&= \left\{ \begin{array}{cc} \varphi((t_1, t_2, \dots, t_r)(t_j)) & \text{if } x = \varphi(t_j) \\ \varphi(\varphi^{-1}(x)) & \text{if } x \neq \varphi(t_j) \ \forall\, j = 1, \dots r \end{array} \right\} \\
&= \left\{ \begin{array}{cc} \varphi(t_{j+1} \text{ or } t_1 \text{ if } j = r)) & \text{if } x = \varphi(t_j) \\ x & \text{if } x \neq \varphi(t_j) \ \forall\, j = 1, \dots r \end{array} \right\} \\
&= (\varphi(t_1), \varphi(t_2), \dots, \varphi(t_r))(x) \ ,
\end{aligned}$$

i.e., $\varphi \circ ((t_1, t_2, \dots, t_r) \circ \varphi^{-1} = (\varphi(t_1), \varphi(t_2), \dots, \varphi(t_r))$. Thus, the conjugate of an $r$-cycle is another $r$-cycle. Moreover, if two cycles are disjoint, then their conjugates by $\varphi$ are also disjoint, because $\varphi$ is 1-1. Therefore, if $\psi$ in $\mathcal{S}_n$ has disjoint cycle decomposition

$$\psi = \gamma_1 \gamma_2 \dots \gamma_m \ ,$$

then

$$\varphi\psi\varphi^{-1} = (\varphi\gamma_1\varphi^{-1})(\varphi\gamma_2\varphi^{-1}) \dots (\varphi\gamma_m\varphi^{-1}) \ ;$$

the last expression is the disjoint cycle decomposition of the conjugate $\varphi\psi\varphi^{-1}$ of $\psi$, and it has the same number of 2-cycles, the same number of 3-cycles, the same number of 4-cycles, and so on

as does the disjoint cycle decomposition of $\psi$. So if two elements of $\mathcal{S}_n$ are conjugate, then their disjoint cycle decompositions have the same "form", i.e., the same number of cycles of each length.

We claim that the converse is also true: If two elements $\psi$ and $\rho$ are such that their disjoint cycle decompositions have the same number of cycles of each length, then we can find a $\varphi$ for which $\rho = \varphi \circ \psi \circ \varphi^{-1}$. This is probably easier to see with an example rather than trying to write it out in symbols: In $\mathcal{S}_{10}$, take

$$\psi = (1,2,3)(6,8,10)(4,9) \qquad \text{and} \qquad \rho = (3,1,5)(7,2,4)(9,10) \ ,$$

two elements of the same "form", in the sense above. One choice for $\varphi$ would be

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 1 & 5 & 9 & ? & 7 & ? & 2 & 10 & 4 \end{pmatrix} \ ,$$

where the two question marks are 6 and 8 in either order. But because we could also write the same $\rho$ in a different way:

$$\rho = (1,5,3)(2,4,7)(10,9) \ ,$$

another choice for $\varphi$ would be

$$\varphi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 5 & 3 & 10 & ? & 2 & ? & 4 & 9 & 7 \end{pmatrix} \ ,$$

where the question marks are 6 and 8 in either order. At any rate, we have shown that two elements of $\mathcal{S}_n$ are conjugate (in $\mathcal{S}_n$) if and only if their disjoint cycle decompositions have the same number of cycles of each length.

The other context in which we can give a different interpretation of conjugacy is in $GL(n,\mathbb{R})$; but for this one we need to recall almost all of our linear algebra course, just to get the notation straight. Recall that, given an $m$-dimensional (real) vector space $V$ and an ordered basis $B = \{\vec{v}_1, \ldots, \vec{v}_m\}$ of $V$, we get a simple function from $[\cdot]_B : V \to \mathbb{R}^n$, given by, if $\vec{v}$ in $V$ is written in terms of $B$ as

$$\vec{v} = b_1\vec{v}_1 + \cdots + b_m\vec{v}_m \ ,$$

then the *coordinate vector* of $\vec{v}$ with respect to $B$ is

$$[\vec{v}]_B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \ .$$

Also, if $W$ is an $n$-dimensional vector space with ordered basis $C = \{\vec{w}_1, \ldots, \vec{w}_n\}$, and if $T : W \to V$ is a linear transformation, then we write the images under $T$ of the elements of $C$ in terms of $B$:

$$\begin{aligned} T(\vec{w}_1) &= t_{1,1}\vec{v}_1 + \cdots + t_{m,1}\vec{v}_m \\ \vdots \quad & \quad \vdots \quad \ddots \quad \vdots \\ T(\vec{w}_n) &= t_{1,n}\vec{v}_1 + \cdots + t_{m,n}\vec{v}_m \end{aligned}$$

Now we can isolate the coefficients $t_{i,j}$ in these equations, make a matrix out of them, and get the *matrix representation* of $T$ with respect to the bases $C$ and $B$:

$$[T]_B^C = \begin{bmatrix} t_{1,1} & \cdots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{m,1} & \cdots & t_{m,n} \end{bmatrix} = \begin{bmatrix} [T(\vec{w}_1)]_B & \cdots & [T(\vec{w}_n)]_B \end{bmatrix} .$$

The reason for doing this is that it converts the abstract linear transformation $T$ into matrix multiplication: If we want the coordinate vector with respect to $B$ of the image under $T$ of any vector $\vec{w}$ in $W$, we first find the coordinate vector of $\vec{w}$ with respect to $C$ and multiply it by the matrix representation of $T$: If

$$\vec{w} = c_1\vec{w}_1 + \cdots + c_n\vec{w}_n \ ,$$

so that

$$[\vec{w}]_C = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \ ,$$

then

$$\begin{aligned} T(\vec{w}) &= c_1T(\vec{w}_1) + \cdots + c_nT(\vec{w}_n) \\ &= c_1(t_{1,1}\vec{v}_1 + \cdots + t_{m,1}\vec{v}_m) + \cdots + c_n(t_{1,n}\vec{v}_1 + \cdots + t_{m,n}\vec{v}_m) \\ &= (c_1t_{1,1} + \cdots + c_nt_{1,n})\vec{v}_1 + \cdots + (c_1t_{m,1} + \cdots + c_nt_{m,n})\vec{v}_m \\ &= \left(\sum_{j=1}^{n} t_{1,j}c_j\right)\vec{v}_1 + \cdots + \left(\sum_{j=1}^{n} t_{m,j}c_j\right)\vec{v}_m \ , \end{aligned}$$

so that we have

$$[T(\vec{w})]_B = \begin{bmatrix} \sum_{j=1}^{n} t_{1,j}c_j \\ \vdots \\ \sum_{j=1}^{n} t_{m,j}c_j \end{bmatrix} = \begin{bmatrix} t_{1,1} & \dots & t_{1,n} \\ \vdots & \ddots & \vdots \\ t_{m,1} & \dots & t_{m,n} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = [T]_B^C[\vec{w}]_C \ .$$

The ends of that last equation give the equation we are shooting for. Notice that what it says is that we can "cancel" the basis $C$ from "top" and "bottom" (which is why I chose this notation). Moreover, if $U$ is still another vector space, with ordered basis $A$, and if $S : U \to W$ is a linear transformation, then for every $\vec{u}$ in $U$ we have

$$[TS]_B^A[\vec{u}]_A = [TS(\vec{u})]_B = [T]_B^C[S(\vec{u})]_C = [T]_B^C[S]_C^A[\vec{u}]_A \ ,$$

so $[TS]_B^A = [T]_B^C[S]_C^A$; i.e., the "cancellation of bases" stuff works on matrices as a whole, and not just on one matrix and and one (coordinate) vector.

One last point before we leave the context of abstract vector spaces: Suppose $n = m$ and $V = W$, so that $C$ and $B$ are two bases for the same vector space. Then to get from the coordinate vector with respect to $C$ of some element $\vec{v}$ of $V$ to the coordinate vector of that $\vec{v}$ with respect to $B$, we can apply the equation above with the identity function $id$ on $V$ — which <u>is</u> a linear transformation from $V$ to itself:

$$[\vec{v}]_B = [id(\vec{v})]_B = [id]_B^C[\vec{v}]_C \ .$$

So $[id]_B^C$ is the *transition matrix* from (the coordinate vectors with respect to) $C$ to (the ones with respect to) $B$. (Notice that the columns of $[id]_B^C$ are just the coordinate vectors of the elements of $C$ with respect to $B$.) Because we have

$$[id]_B^C[id]_C^B = [id]_B^B = I \text{ (the identity matrix)} \qquad \text{and} \qquad [id]_C^B[id]_B^C = I \ ,$$

these transition matrices are inverses of each other.

So what has all this got to do with conjugation in $GL(n,\mathbb{R})$? Well, suppose we restrict to $n = m$ and $V = W = \mathbb{R}^n$. Then for any $n \times n$ matrix $M$, multiplication by $M$ gives a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, and $M$ is the matrix representation of $T$ with respect to the standard basis

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} .$$

Now take any element $P$ of $GL(n,\mathbb{R})$, i.e., an invertible $n \times n$ matrix. Then the columns of $P$ form a basis $C$ for $\mathbb{R}^n$, and those columns are the coordinate vectors of themselves with respect to the standard basis $B$; so $P = [id]_B^C$. Thus, $P^{-1} = [id]_C^B$, and we have

$$P^{-1}MP = [id]_C^B[T]_B^B[id]_B^C = [T]_C^C .$$

In other words, a conjugate of $M$ is just the matrix representation of multiplication by $M$, regarded as a linear transformation on $\mathbb{R}^n$, but with respect to a possibly different basis. (This works even if $M$ isn't invertible, i.e., if $M \notin GL(n,\mathbb{R})$.)