The general linear group is a beautiful butterfly

Joshua Ruiter

September 27, 2021

The goal of this expository paper is to describe an analogy, due to Alexander Grothendieck, between the general linear group $\mathrm{GL}_n(\mathbb{R})$ and a pinned butterfly specimen. We describe how different structural features of $\mathrm{GL}_n(\mathbb{R})$ correspond to various parts of the butterfly, and describe several equations relating these parts of $\mathrm{GL}_n(\mathbb{R})$ to each other.

We assume the reader is familiar with the real numbers $\mathbb{R}$, matrix multiplication, and writing functions as $f : X \to Y$. It is useful to know some basic group theory, but not strictly necessary.

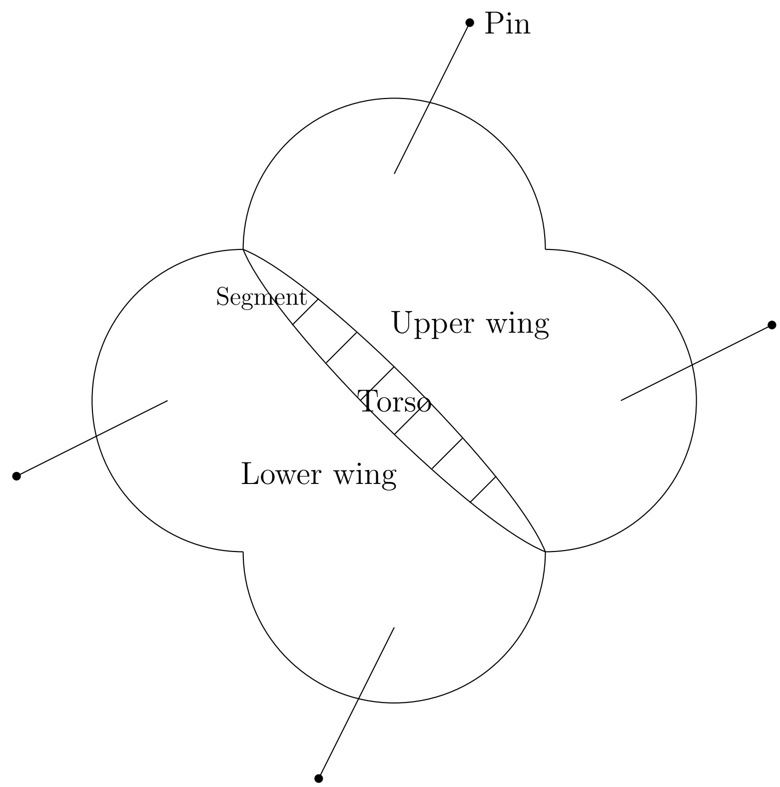
Contents

1 Structural features of $\mathrm{GL}_n(\mathbb{R})$

Let n be a positive integer. The **general linear group** $\mathrm{GL}_n(\mathbb{R})$ is the group of $n \times n$ invertible matrices, a.k.a. matrices with determinant not equal to zero. In this context, "group" is shorthand for the fact that the product of two invertible matrices is invertible, and the inverse of an invertible matrix is invertible.

Example 1.1. When $n = 1$, we have 1×1 matrices with nonzero determinant, so the only requirement is that the single entry must be nonzero. In other words, we can identify $\mathrm{GL}_1(\mathbb{R})$ with the set of nonzero real numbers $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$.

The goal of this paper is to make an analogy between $\mathrm{GL}_n(\mathbb{R})$ and a pinned butterfly specimen, something you might see in a museum. This analogy comes from the general theory of reductive algebraic groups, but we'll just study $\mathrm{GL}_n(\mathbb{R})$, and you don't need to know what "reductive" or "algebraic group" mean to understand what's going on. Here is a diagram with all the parts of the butterfly which we will eventually identify with various aspects of $\mathrm{GL}_n(\mathbb{R})$.



In this picture, we have a main torso with segments along it, two wings, and some pins holding the wings in place. Each of these elements of the diagram corresponds to some feature of $\mathrm{GL}_n(\mathbb{R})$, and we will revisit this diagram each time as we develop the analogy. We will also eventually describe a feature of $\mathrm{GL}_n(\mathbb{R})$ corresponding to the DNA of the butterfly, though this will come last.

1.1 Diagonal subgroup

Definition 1.2. The **diagonal subgroup** $D \subset \mathrm{GL}_n(\mathbb{R})$ is the subgroup of diagonal matrices with nonzero entries.

$$D = \left\{ \begin{pmatrix} d_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & d_n \end{pmatrix} : d_1, \dots, d_n \in \mathbb{R}^\times \right\}$$

The diagonal entries have to be nonzero otherwise the determinant would be zero. (Remember: the determinant of a diagonal matrix is the product of the diagonal entries.)

Notation 1.3. From now on, when writing matrices with zeros, we omit the zeros and just leave those spaces blank. For example, in this notation D is

$$D = \left\{ \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} : d_1, \dots, d_n \in \mathbb{R}^\times \right\}$$

For brevity, we can also denote the diagonal matrix with entries $d_1, \dots, d_n$ by $\mathrm{diag}(d_1, \dots, d_n)$. One important diagonal matrix is the identity matrix.

$$I_n = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}$$

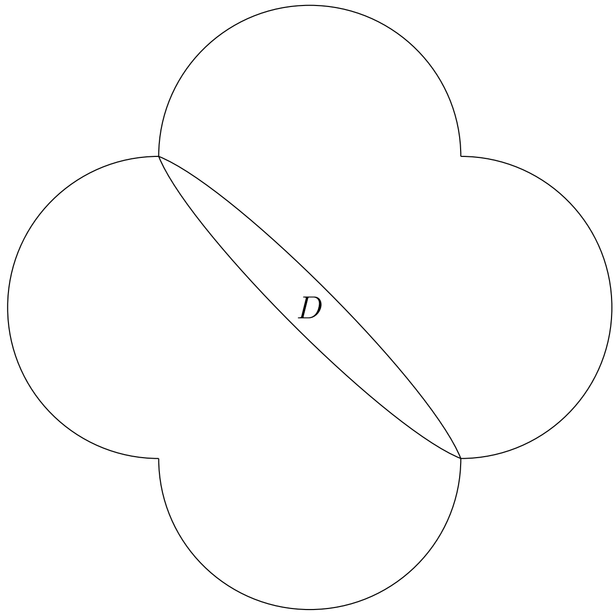
This is called the identity matrix because it has the property that for any $n \times n$ matrix A, $AI_n = I_nA = A$. The fact that D is a subgroup of $\mathrm{GL}_n(\mathbb{R})$ is a sophisticated way of saying that the product of two diagonal matrices is diagonal, and the inverse of a diagonal matrix is diagonal. Multiplying diagonal matrices together is very quick - just multiply the corresponding entries. For example, the product of two 2×2 diagonal matrices is given by

$$\begin{pmatrix} a_1 & \\ & a_2 \end{pmatrix} \begin{pmatrix} b_1 & \\ & b_2 \end{pmatrix} = \begin{pmatrix} a_1b_1 & \\ & a_2b_2 \end{pmatrix}$$

A consequence of this is that to invert a diagonal matrix, all we have to do is invert each diagonal entry. Usually computing matrix inverses is hard, but not with diagonal matrices.

$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}^{-1} = \begin{pmatrix} d_1^{-1} & & \\ & \ddots & \\ & & d_n^{-1} \end{pmatrix}$$

In our butterfly analogy, D is the torso of the butterfly. This is why we draw our butterfly "at an angle" in our diagram, so that the torso goes diagonally from top left to bottom right, just like D does inside of $\mathrm{GL}_n(\mathbb{R})$.

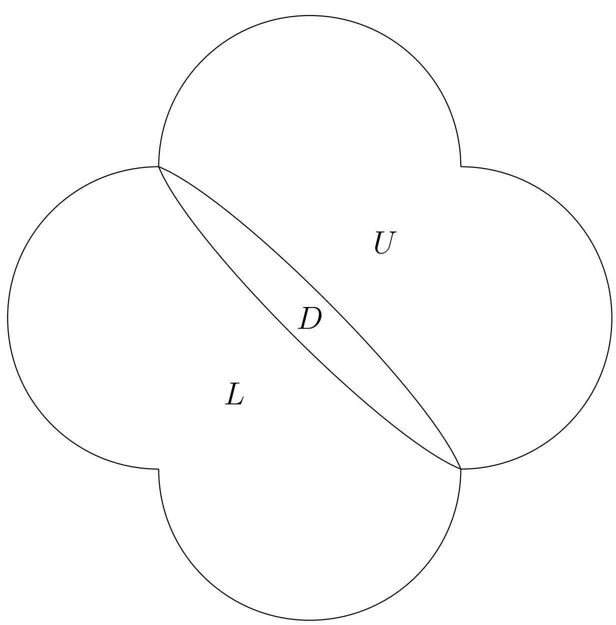


Definition 1.4. Let $U \subset G$ be the subgroup of upper triangular matrices. Similarly, L will be the lower triangular matrices. For example, when $n = 3$, these are matrices of the following shapes.

$$U = \left\{ \begin{pmatrix} * & * & * \\ & * & * \\ & & * \end{pmatrix} \right\} \qquad L = \left\{ \begin{pmatrix} * & & \\ * & * & \\ * & * & * \end{pmatrix} \right\}$$

In U, the entries above the main diagonal can be anything, but the entries along the diagonal have to be nonzero (the determinant of an upper triangular matrix is the product of the diagonal entries). Similarly, in L the entries below the main diagonal can be anything, but entries along the diagonal must be nonzero.

In terms of the butterfly analogy, U and L are the wings of the butterfly. They reach out in opposite directions from the diagonal D.



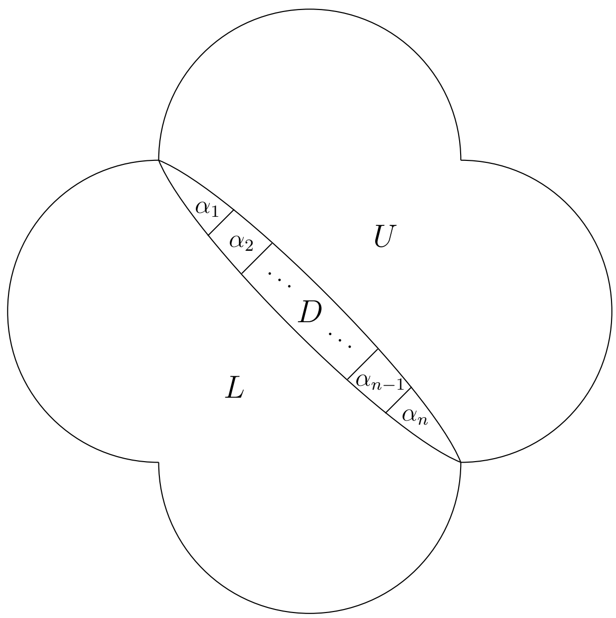
1.2 Character functions

Next we describe the part of $\mathrm{GL}_n(\mathbb{R})$ which corresponds to segments along the torso of the butterfly.

Definition 1.5. For $i = 1, \ldots, n$, define the ***i*th character function** of D to be the function which sends a diagonal matrix to its ith diagonal entry. I will denote this function by α_i.

$$\alpha_i : D \to \mathbb{R}^\times \qquad \alpha_i \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = d_i$$

In our analogy, the character functions are the separate segments of the torso. There are n characters functions, so in our butterfly diagram there should be n segments.



Example 1.6. Here are some examples of characters evaluated in the case $n = 3$.

$$\alpha_1 \begin{pmatrix} 3 & & \\ & -2 & \\ & & 5 \end{pmatrix} = 3 \qquad \alpha_2 \begin{pmatrix} 3 & & \\ & -2 & \\ & & 5 \end{pmatrix} = -2$$

Definition 1.7. Let $d \in D$ be a diagonal matrix. Given two character functions $\alpha_i, \alpha_j : D \to \mathbb{R}^\times$, their sum, negation, and difference are defined as follows.

$$\begin{aligned} (\alpha_i + \alpha_j)(d) &= \alpha_i(d) \cdot \alpha_j(d) = d_i d_j \\ (-\alpha_j)(d) &= \alpha_j(d)^{-1} = d_j^{-1} \\ (\alpha_i - \alpha_j)(d) &= (\alpha_i + (-\alpha_j))(d) = \alpha_i(d)\alpha_j(d)^{-1} = d_i d_j^{-1} \end{aligned}$$

Example 1.8. For example,

$$(\alpha_1 + \alpha_2) \begin{pmatrix} 3 & & \\ & -2 & \\ & & 5 \end{pmatrix} = -6 \qquad (\alpha_1 - \alpha_2) \begin{pmatrix} 3 & & \\ & -2 & \\ & & 5 \end{pmatrix} = -\frac{3}{2}$$

More generally, inside $\mathrm{GL}_2(\mathbb{R})$ with the character functions α_1, α_2 and the diagonal matrix $d = \mathrm{diag}(d_1, d_2) \in D$.

$$\alpha_1 \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix} = d_1 \qquad \alpha_2 \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix} = d_2$$

The sum $\alpha_1 + \alpha_2$ is another function $D \to \mathbb{R}^\times$, which is the function

$$(\alpha_1 + \alpha_2) \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix} = \alpha_1(d) \cdot \alpha_2(d) = d_1 d_2$$

And the difference is

$$(\alpha_1 - \alpha_2)(d) = d_1 d_2^{-1}$$

Remark 1.9. Why not define addition of characters "more logically" as below?

$$(\alpha_i + \alpha_j)(d) \stackrel{?}{=} \alpha_i(d) + \alpha_j(d)$$

The problem is that with this definition, the function $\alpha_i + \alpha_j$ no longer is a function to $\mathbb{R}^\times$. For example, if d is the matrix

$$\begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$$

then $(\alpha_1 + \alpha_2)(d) = 0$, which is a problem. Since I want the sum of two character functions to also be a function $D \to \mathbb{R}^\times$, I have to define it in the way I did.

1.3 Elementary matrices and pinning functions

Definition 1.10. For each pair (i, j) with $1 \leq i, j \leq n$ and $i \neq j$, and a real number x, the **elementary matrix** $e_{ij}(x)$ is the matrix with x in the ijth entry, 1's on the diagonal, and zeros everywhere else. For example, inside $\mathrm{GL}_2(\mathbb{R})$,

$$e_{12}(x) = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \qquad e_{21}(x) = \begin{pmatrix} 1 & \\ x & 1 \end{pmatrix}$$

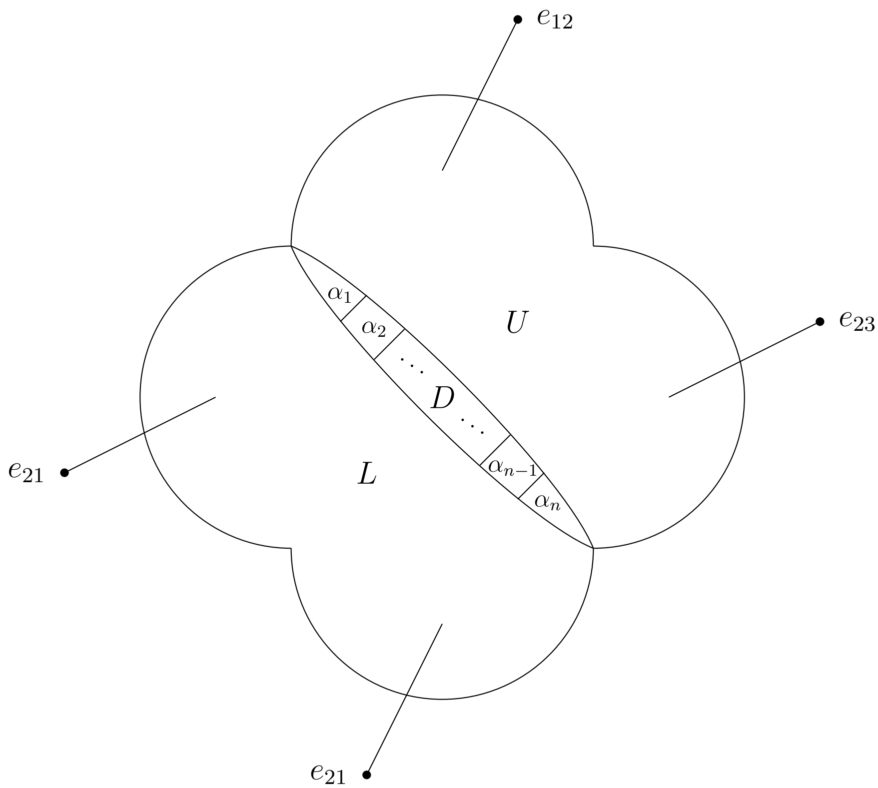
And inside $\mathrm{GL}_3(\mathbb{R})$,

$$e_{12}(x) = \begin{pmatrix} 1 & x & \\ & 1 & \\ & & 1 \end{pmatrix} \qquad e_{13}(x) = \begin{pmatrix} 1 & & x \\ & 1 & \\ & & 1 \end{pmatrix}$$

There is theoretically potential for confusion between these two different meanings of $e_{12}(x)$, but practically speaking we almost always have some fixed n in mind. And anyway, the two matrices both called $e_{12}(x)$ aren't that different anyway, the only difference is an extra row and column with just a 1 on the diagonal and zero elsewhere.

Instead of just thinking of $e_{ij}(x)$ as an individual matrix, think of $e_{ij}(-)$ as a function. It is a function $e_{ij} : \mathbb{R} \to \mathrm{GL}_n(\mathbb{R})$, whose input is x, and output is $e_{ij}(x)$. Note that $e_{ij}(0) = I_n$ for any i and j.

In terms of the butterfly analogy, each function $e_{ij} : \mathbb{R} \to \mathrm{GL}_n(\mathbb{R})$ is a pin. There are $n^2 - n$ of them, and together, they hold the wings (U and L) fixed in place. Because of this analogy the functions e_{ij} are called **pinning functions**.



Notice that e_{12} and e_{23} are depicted pinning the upper triangular subgroup as $e_{12}(x)$ and $e_{23}(x)$ are upper triangular matrices (for any x), while $e_{21}(x)$ and $e_{32}(x)$ are lower triangular so they are pinning L in the picture.

Remark 1.11. Elementary matrices encode the most useful elementary row operation involved in Gaussian elimination, namely adding a multiple of one row to another row. Specifically, multiplying an $n \times n$ matrix A by $e_{ij}(x)$ on the left modifies A by adding x times the jth row to the ith row. As an equation,

$$e_{ij}(x) \cdot A = A'$$

where A' is the matrix obtained by adding x times the jth row of A to the ith row of A. For concreteness, let's suppose $i = 1, j = 2$ and A is 2×2.

$$e_{12}(x) \cdot A = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + xa_{21} & a_{12} + xa_{22} \\ a_{21} & a_{22} \end{pmatrix}$$

Similarly, multiplying A by $e_{ij}(x)$ on the right modifies A by a column operation, but column operations are not important for the rest of our discussion. On the other hand, our ability to conceptualize $e_{ij}(x)$ as a column operation will be critical to some proofs later.

This concludes our description of the structural features of $\mathrm{GL}_n(\mathbb{R})$ corresponding to parts of a butterfly. As already alluded to there is one more analogy to make with the DNA of the butterfly, but we aren't quite ready yet. Here is a table summarizing the analogy so far.

Group	Butterfly
$\mathrm{GL}_n(\mathbb{R})$	Whole butterfly
D	Torso
U, L	Wings
α_i	Segment on torso
$e_{ij}(-)$	Pin

2 Formulas

Now that we've set the stage with the various parts of $\mathrm{GL}_n(\mathbb{R})$ corresponding to parts of our metaphorical butterfly, we can start talking about the relationships between these components.

2.1 Exponential formula

Our first formula concerns the pinning maps.

Proposition 2.1 (Exponential formula)**.** *For any $i \neq j$ and any $x, y \in \mathbb{R}$, $e_{ij}(x+y) = e_{ij}(x) \cdot e_{ij}(y)$.*

Example 2.2. Before the proof, let's check this in a particular case, e_{12} inside $\mathrm{GL}_2(\mathbb{R})$.

$$e_{12}(x) \cdot e_{12}(y) = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} 1 & y \\ & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+y \\ & 1 \end{pmatrix} = e_{12}(x+y)$$

Proof. Recall that we can think of $e_{ij}(x)$ in terms of row operations. As a row operation, $e_{ij}(x+y)$ means I add $(x+y)$ times the jth row to the ith row. On the other hand, $e_{ij}(x) \cdot e_{ij}(y)$ (multiplied on the left of some matrix A) means first add y times the jth row to the ith row, then add x times the jth row to the ith row. But obviously that's the same as just doing $(x+y)$ times the jth row to the ith row all in one step. □

In algebraic terms, the exponential formula says that the map $e_{ij} : \mathbb{R} \to \mathrm{GL}_n(\mathbb{R})$ is a group homomorphism, if we view $\mathbb{R}$ as a group under addition and $\mathrm{GL}_n(\mathbb{R})$ as a group under matrix multiplication.

Corollary 2.3. *The inverse of the matrix $e_{ij}(x)$ is $e_{ij}(-x)$.*

Proof. Recall that the inverse of a matrix A is, by definition, the matrix A^{-1} with the property $AA^{-1} = I_n$. We can check then that

$$e_{ij}(x) \cdot e_{ij}(-x) = e_{ij}(x + (-x)) = e_{ij}(0) = I_n$$

□

Remark 2.4. We have called this the "exponential formula" because it looks like an exponent rule that you already know. Remember that

$$e^{x+y} = e^x e^y$$

(Actually, you can replace e with any number here.) The equation

$$e_{ij}(x+y) = e_{ij}(x) \cdot e_{ij}(y)$$

is analogous to this. If you study far enough into the theory of Lie groups and Lie algebras, or algebraic groups (and their Lie algebras), then the analogy here becomes more precise. These maps e_{ij} are in fact generalizations of the exponential map, but this involves much more theory than we have space to discuss.

2.2 Conjugation formula and root system

The next formula describes how a pinning function interacts with the diagonal subgroup D and the character functions. It involves a **conjugation** computation. We're going to take a diagonal matrix d and an elementary matrix $e_{ij}(x)$, and compute

$$d \cdot e_{ij}(x) \cdot d^{-1}$$

Remember that in general, matrix multiplication is not commutative. If it was, then $d \cdot e_{ij}(x) = e_{ij}(x) \cdot d$ and the expression above would just be equal to $e_{ij}(x)$ because the d and d^{-1} would cancel out. But these matrices do not commute (meaning $d \cdot e_{ij}(x) \neq e_{ij}(x) \cdot d$), so we shouldn't expect to get $e_{ij}(x)$ at the end. However, we will get a matrix that looks like $e_{ij}(x)$, with a different input than x.

Proposition 2.5 (Conjugation formula, incomplete)**.** *For any $i \neq j$, any $d \in D$, and any $x \in \mathbb{R}$,*

$$d \cdot e_{ij}(x) \cdot d^{-1} = e_{ij}(\qquad)$$

To figure out what goes in the blank, we'll work out an example with $i = 1$ and $j = 2$ and stay inside $\mathrm{GL}_2(\mathbb{R})$, but let x be any real number, and d any (invertible) diagonal matrix. First invert d by inverting each diagonal entry. Then do the matrix multiplication.

$$\begin{aligned} d \cdot e_{12}(x) \cdot d^{-1} &= \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \begin{pmatrix} d_1^{-1} & \\ & d_2^{-1} \end{pmatrix} \\ &= \begin{pmatrix} d_1 & d_1 x \\ & d_2 \end{pmatrix} \begin{pmatrix} d_1^{-1} & \\ & d_2^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 1 & d_1 x d_2^{-1} \\ & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & d_1 d_2^{-1} x \\ & 1 \end{pmatrix} \\ &= e_{12}(d_1 d_2^{-1} x) \end{aligned}$$

In one step, I reversed the order of x and d_2^{-1}. Why is this ok, but it wasn't ok to do this earlier? Earlier it was with matrices, which isn't valid, but now this is with real numbers, where such a reversal is valid. We still aren't done. We did the calculation, but I want to rewrite the final result, because the expression $d_1 d_2^{-1}$ should look familiar. Using the fact that

$$(\alpha_1 - \alpha_2)(d) = d_1 d_2^{-1}$$

I can rewrite it again as

$$d \cdot e_{12}(x) \cdot d^{-1} = e_{12}(d_1 d_2^{-1} x) = e_{12}\Big((\alpha_1 - \alpha_2)(d) \cdot x\Big)$$

The general statement just replaces 1 with i and 2 with j.

Proposition 2.6 (Conjugation formula)**.** *For any $i \neq j$, any $d \in D$, and any $x \in \mathbb{R}$,*

$$d \cdot e_{ij}(x) \cdot d^{-1} = e_{ij}\Big((\alpha_i - \alpha_j)(d) \cdot x\Big)$$

Proof. Again we exploit the fact that $e_{ij}(x)$ represents an elementary row operation. When we multiply d^{-1} on the left by $e_{ij}(x)$, we add x times the jth row (which is just a d_j^{-1} in the jth column and zeros elsewhere) to the ith row (which has a zero in the jth spot, because d^{-1} is diagonal.) So we end up with d^{-1}, except with a d_j^{-1} in the ijth spot.

$$e_{ij}(x) \cdot d^{-1} = \begin{pmatrix} d_1^{-1} & & & & & & & \\ & d_2^{-1} & & & & & & \\ & & \ddots & & & & & \\ & & & d_i^{-1} & \cdots & d_j^{-1} & & \\ & & & & \ddots & \vdots & & \\ & & & & & d_j^{-1} & & \\ & & & & & & \ddots & \\ & & & & & & & d_n^{-1} \end{pmatrix}$$

Now we multiply this by d (on the left). We aren't quite multiplying two diagonal matrices, but what ends up happening is that we multiply the corresponding diagonal entries, and this one nonzero off-diagonal entry of $e_{ij}(x) \cdot d^{-1}$ gets multiplied by d_i because it is in the

ith row. So we get

$$d \cdot e_{ij}(x) \cdot d^{-1} = \begin{pmatrix} d_1 d_1^{-1} & & & & & & \\ & \ddots & & & & & \\ & & d_i d_i^{-1} & \cdots & d_i d_j^{-1} & & \\ & & & \ddots & \vdots & & \\ & & & & d_j d_j^{-1} & & \\ & & & & & \ddots & \\ & & & & & & d_n d_n^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & d_i d_j^{-1} & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

$$= e_{ij}\left(d_i d_j^{-1} x\right) = e_{ij}\Big((\alpha_i - \alpha_j)(d) \cdot x\Big)$$

□

How should we think about this equation in terms of the butterfly analogy? It involves several of the pieces of $\mathrm{GL}_n(\mathbb{R})$ as a butterfly that we've talked about, and relates them together in a single equation. It relates the torso (the diagonal subgroup D where d comes from), the segments of the torso (characters α_i, α_j), and a single pin (e_{ij}).

Definition 2.7. Because of their important role in the conjugation formula, the differences of characters $\alpha_i - \alpha_j$ have a name. The difference $\alpha_i - \alpha_j$ is called a **root**, and the set of all these is the **root system** Φ of $\mathrm{GL}_n(\mathbb{R})$.

$$\Phi = \{\alpha_i - \alpha_j : i \neq j, 1 \leq i \leq n, 1 \leq j \leq n\}$$

Φ is not a group or vector space or anything special, it is just a set for now. To make notation easier, we will define

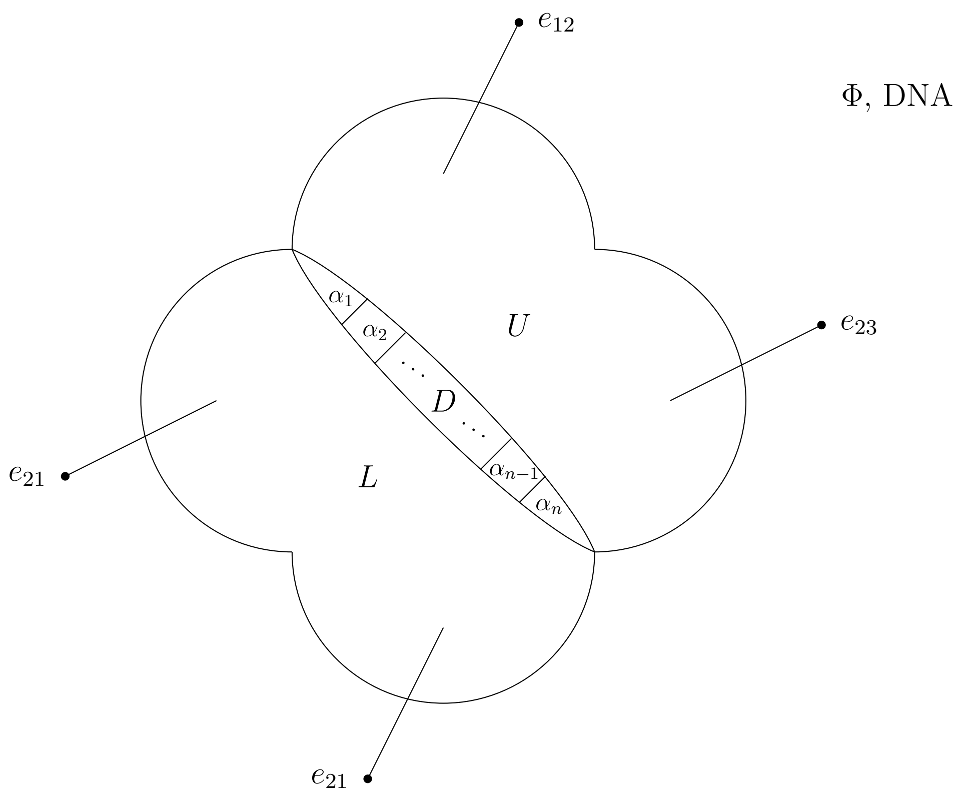
$$\alpha_{ij} = \alpha_i - \alpha_j$$

Notice that there is one root for each pinning function, because for each (i, j) with $1 \leq i, j \leq n$ and $i \neq j$, we have a pinning function e_{ij} and a root α_{ij}. So just as there are $n^2 - n$ pinning functions, there are $n^2 - n$ roots.

Remark 2.8. Using the notation $\alpha_{ij} = \alpha_i - \alpha_j$, we can rewrite the conjugation formula more compactly as

$$d \cdot e_{ij}(x) \cdot d^{-1} = e_{ij}\Big(\alpha_{ij}(d)x\Big)$$

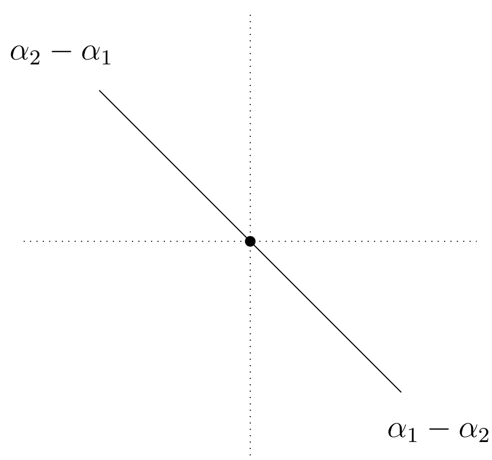
In terms of the butterfly analogy, the root system is the DNA. It's invisible to the eye, but it pulls the strings. Just like a biologist wouldn't study a butterfly without considering DNA, a mathematician studying $\mathrm{GL}_n(\mathbb{R})$ should keep Φ in mind.



Setting aside $\mathrm{GL}_n(\mathbb{R})$ and the butterfly analogy for the moment, let's draw a more geometric picture of Φ, starting with the case $n = 2$, where there are just two roots.

$$\Phi = \{\alpha_1 - \alpha_2, \alpha_2 - \alpha_1\}$$

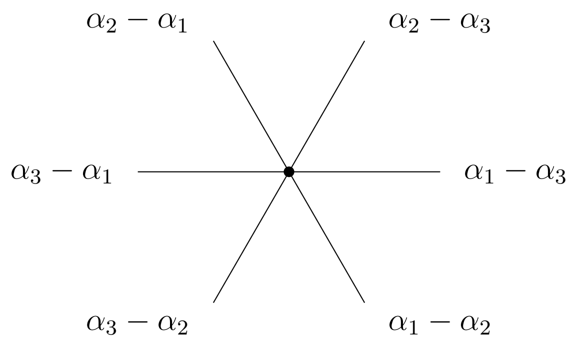
To draw this, I'll draw an xy-plane with the x-axis labelled α_1, and the y-axis labelled α_2.



This is called the A_1 root system. Now let's draw the case $n = 3$, where there are 6 roots.

$$\Phi = \{\alpha_1 - \alpha_2, \alpha_2 - \alpha_1, \alpha_1 - \alpha_3, \alpha_3 - \alpha_1, \alpha_2 - \alpha_3, \alpha_3 - \alpha_2\}$$

If I try to draw this like the other one, I would need to draw it in three dimensions using α_1 as x-axis, α_2 as y-axis, and α_3 as z-axis. That's kind of hard, though, so I want to do something easier. Instead, I'm going to start out by observing that all of these are "perpendicular" to the vector $\alpha_1 + \alpha_2 + \alpha_3$. In other words, if I draw those 6 vectors above in $\mathbb{R}^3$, all of them lie in the plane perpendicular to the vector $\langle 1, 1, 1 \rangle$. So I'll just draw the picture inside that plane, because then I only need two dimensions (instead of three).



This is called the A_2 root system. Notice that $\alpha_i - \alpha_j$ is opposite from $\alpha_j - \alpha_i$ for every pair (i, j). Drawing the $n = 4$ or higher cases get into more dimensions and gets a lot harder. Wikipedia has a picture of the $n = 4$ case (which is called type A_3) on their page on root systems: `https://en.wikipedia.org/wiki/Root_system#/media/File:A3vzome.jpg`. In general, the root system of $\mathrm{GL}_n(\mathbb{R})$ is the type A_{n-1} root system.

Remark 2.9. Four quick observations about our picture of A_2.

1. Φ is a finite set of vectors not containing the zero vector. (Because $i = j$ is not allowed for a root $\alpha_i - \alpha_j$.)

2. For any $v \in \Phi$, the scalar multiples of v which are in Φ are exactly v and $-v$. That is, no scalar multiples other than $\pm v$ are included in Φ, and $-v$ is always included if v is. (Concretely, $-\alpha_{ij} = \alpha_{ji} \in \Phi$.)

3. For any $v \in \Phi$, if you take the line perpendicular to v and reflect Φ across that line, you get Φ back. (Try some examples to convince yourself.)

4. For any $v, w \in \Phi$, the dot product $v \cdot w$ is an integer. [1]

These are the four axioms to have a root system. A root system is a very nice symmetrical structure, and one showing up here when talking about $\mathrm{GL}_n(\mathbb{R})$ is no coincidence. This is part of a really big pattern and theorem which classifies algebraic groups in terms of root systems.

[1] To take dot products of roots, use the rule

$$\alpha_i \cdot \alpha_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

For example,

$$\alpha_{12} \cdot \alpha_{23} = (\alpha_1 - \alpha_2) \cdot (\alpha_2 - \alpha_3) = \alpha_1 \cdot \alpha_2 - \alpha_1 \cdot \alpha_3 - \alpha_2 \cdot \alpha_2 + \alpha_2 \cdot \alpha_3 = 0 - 0 + 1 - 0 = 1$$

One last thing to notice about Φ is that sometimes you can add two elements of Φ and end up back in Φ, but other times the sum is not in Φ. For example,

$$(\alpha_1 - \alpha_2) + (\alpha_2 - \alpha_3) = \alpha_1 - \alpha_3 \in \Phi$$
$$(\alpha_1 - \alpha_2) + (\alpha_1 - \alpha_3) = 2\alpha_1 - \alpha_2 - \alpha_3 \notin \Phi$$

This aspect of Φ will come into play in our last formula.

2.3 Chevalley commutator formula

There is one last formula, which relates the pinning functions to each other. We'll talk about how they interact in terms of commutators.

Definition 2.10. Let $A, B \in \mathrm{GL}_n(\mathbb{R})$ be invertible matrices. Their **commutator** is

$$[A, B] = ABA^{-1}B^{-1}$$

One way to think of commutators is that $[A, B]$ measure how "far" A, B are from commuting. That is to say,

$$[A, B] = I_n \iff AB = BA$$

So if $[A, B]$ is "as simple as possible" in the sense that it is the identity matrix, then A and B commute. But if $[A, B]$ is more complicated (i.e. not the identity) then A and B don't commute. Yet another way to rearrange this is to say that

$$AB = [A, B]BA$$

which roughly expresses the same idea, that $[A, B]$ is the "defect" of A and B failing to commute (if $[A, B]$ is not the identity).

Definition 2.11. To emphasize the connection between the root α_{ij} and the pinning map e_{ij}, we'll write $e_{\alpha_{ij}}$ instead of e_{ij}. It's still the same function, it just has a slightly different name now.

With this notation, we can state the final formula, which is named after French mathematician Claude Chevalley.

Proposition 2.12 (Chevalley commutator formula)**.** *Let i, j, k, ℓ be positive integers and $x, y \in \mathbb{R}$, and suppose $\alpha_{ij} \neq -\alpha_{k\ell}$. Then*

$$[e_{\alpha_{ij}}(x), e_{\alpha_{k\ell}}(y)] = \begin{cases} I_n & \alpha_{ij} + \alpha_{k\ell} \notin \Phi \\ e_{\alpha_{ij}+\alpha_{k\ell}}(\epsilon xy) & \alpha_{ij} + \alpha_{k\ell} \in \Phi \end{cases}$$

where $\epsilon = \pm 1$ depends on i, j, k, ℓ (but not on x, y).

We won't go into the proof; we'll just try to conceptualize what's going on in this equation.

Conceptually, what this equation says is that the additive structure of Φ determines when two elementary matrices commute. That last sentence can be very hard to wrap your head around fully, so let's be more specific. The formula says that if the sum of two roots is not a root and not zero, then the commutator is the identity, or in other words, those elementary matrices commute. But if the sum of two roots is a root, then the commutator is a third elementary matrix (different from the identity). The assumption $\alpha_{ij} \neq -\alpha_{k\ell}$ says to ignore what happens when the sum of two roots is zero. If that happens, the commutator is pretty complicated, but the theorem doesn't say anything about that case, and that case isn't so useful anyway.

We can write the formula more concretely, though the more concrete version obfuscates the connection to Φ. Notice that

$$\alpha_{ij} = -\alpha_{k\ell} \iff i = \ell \text{ and } j = k$$

Equivalently,

$$\alpha_{ij} \neq -\alpha_{k\ell} \iff i \neq \ell \text{ or } j \neq k$$

So the assumption $\alpha_{ij} \neq -\alpha_{k\ell}$ in the theorem can be rephrased as assuming that $i \neq \ell$ or $j \neq k$. Under this assumption,

$$\begin{aligned} \alpha_{ij} + \alpha_{k\ell} \notin \Phi &\iff i \neq \ell \text{ and } j \neq k \\ \alpha_{ij} + \alpha_{k\ell} \in \Phi &\iff (i \neq \ell \text{ and } j = k) \text{ or } (i = \ell \text{ and} j \neq k) \end{aligned}$$

So we can rewrite the commutator formula as

$$[e_{ij}(x), e_{k\ell}(y)] = \begin{cases} I_n & i \neq \ell \text{ and } j \neq k \\ e_{i\ell}(\epsilon xy) & i \neq \ell \text{ and } j = k \text{ (so } \alpha_{ij} + \alpha_{k\ell} = \alpha_{i\ell}) \\ e_{kj}(\epsilon xy) & i = \ell \text{ and } j \neq k \text{ (so } \alpha_{ij} + \alpha_{k\ell} = \alpha_{kj}) \end{cases}$$

In fact, when writing it this way we can actually specify when $\epsilon = 1$ or $\epsilon = -1$, but the sign is not the most important thing here.

$$[e_{ij}(x), e_{k\ell}(y)] = \begin{cases} 1 & i \neq \ell \text{ and } j \neq k \\ e_{i\ell}(xy) & i \neq \ell \text{ and } j = k \text{ (so } \alpha_{ij} + \alpha_{k\ell} = \alpha_{i\ell}) \\ e_{kj}(-xy) & i = \ell \text{ and } j \neq k \text{ (so } \alpha_{ij} + \alpha_{k\ell} = \alpha_{kj}) \end{cases}$$

This version is perhaps easier to understand, but it also hides the fact that Φ is really pulling the strings behind this equation. To try and get a handle on the Chevalley commutator formula, it helps to just do some concrete matrix computations by hand.

Example 2.13. Let $n = 4$, and take the roots α_{12} and α_{34}. Let $x, y \in \mathbb{R}$. Here are the associated elementary matrices.

$$e_{\alpha_{12}}(x) = \begin{pmatrix} 1 & x & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \qquad e_{\alpha_{34}}(y) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & y \\ & & & 1 \end{pmatrix}$$

Does the sum $\alpha_{12} + \alpha_{34}$ belong to Φ? No, it does not. So according to the Chevalley commutator formula,

$$[e_{\alpha_{12}}(x), e_{\alpha_{34}}(y)] = e_{\alpha_{12}}(x) \cdot e_{\alpha_{34}}(y) \cdot e_{\alpha_{12}}(x)^{-1} \cdot e_{\alpha_{34}}(y)^{-1} = I_n$$

or rearranging,

$$e_{\alpha_{12}}(x) \cdot e_{\alpha_{34}}(y) = e_{\alpha_{34}}(y) \cdot e_{\alpha_{12}}(x)$$

So while generally speaking, matrix multiplication is not commutative, it is commutative for these particular matrices. Returning to the full commutator expression involving inverses, remember that by Corollary 2.3, $e_{\alpha_{12}}(x)^{-1} = e_{\alpha_{12}}(-x)$ and $e_{\alpha_{34}}(y) = -y$. So at least the inverses aren't hard to compute. But this is still a potentially complicated product of four matrices.

$$[e_{\alpha_{12}}(x), e_{\alpha_{34}}(y)] = \begin{pmatrix} 1 & x & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & y \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & -x & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 - y & \\ & & & 1 \end{pmatrix}$$

If you want to appreciate the power of the Chevalley commutator formula, just multiply this out by hand. It's not terribly difficult, but you'll see for yourself why this comes out to be the identity matrix and appreciate the fact that in the future you can use this formula to bypass that pen and paper calculation.

Example 2.14. Now let's do an example where the sum is a root. So take α_{12} and α_{23}, and the sum is α_{13}. So the Chevalley commutator formula says that

$$[e_{\alpha_{12}(x)}, e_{\alpha_{23}}(y)] = e_{\alpha_{13}}(\epsilon xy)$$

for some sign $\epsilon = \pm 1$. I encourage you to work out the left hand side for yourself, and verify that in this case $\epsilon = 1$.

2.4 Summary

Let's put everything together. We have our group $\mathrm{GL}_n(\mathbb{R})$, diagonal subgroup D, pinning maps e_{ij}, and root system Φ consisting of roots α_{ij}. Each of these pieces corresponds to a part of a pinned butterfly specimen. We also have three equations relating different elements together. First, the exponential formula tells us something about each individual pin.

$$e_{ij}(x) \cdot e_{ij}(y) = e_{ij}(x + y)$$

Second, the conjugation formula which relates a pinning map e_{ij} with D and the root α_{ij}.

$$d \cdot e_{ij}(x) \cdot d^{-1} = e_{ij}\Big(\alpha_{ij}(d)x\Big)$$

Finally, the Chevalley commutator formula relates two different pinning maps $e_{ij} = e_{\alpha_{ij}}$ and $e_{k\ell} = e_{\alpha_{k\ell}}$ (assuming $\alpha_{ij} \neq -\alpha_{k\ell}$) by expressing their commutator in terms of the root system Φ (in particular, in terms of whether or not the sum $\alpha_{ij} + \alpha_{k\ell}$ belongs to Φ.)

$$[e_{\alpha_{ij}}(x), e_{\alpha_{k\ell}}(y)] = \begin{cases} I_n & \alpha_{ij} + \alpha_{k\ell} \notin \Phi \\ e_{\alpha_{ij}+\alpha_{k\ell}}(\epsilon xy) & \alpha_{ij} + \alpha_{k\ell} \in \Phi \end{cases}$$

The conjugation and commutator formulas are why Φ is the DNA of the butterfly $\mathrm{GL}_n(\mathbb{R})$. Both equations are tightly connected to Φ, but you would not know it from just "looking at the butterfly." That is to say, if you study $\mathrm{GL}_n(\mathbb{R})$ on a surface level, if you look at the left hand side of these equations, it is not at all obvious that Φ is somehow connected to these basic matrix computations. But as it turns out, the root system Φ is deeply connected to these matrix computations.

3 Further explorations

Everything done here for the general linear group $\mathrm{GL}_n(\mathbb{R})$ can be essentially repeated for the special linear group $\mathrm{SL}_n(\mathbb{R})$ (matrices of determinant 1). There is still the diagonal subgroup (now only diagonal matrices of determinant 1), character functions (unchanged), and elementary matrices (unchanged). Since the matrix $e_{ij}(x)$ has determinant 1, all three formulas discussed involve only matrices of determinant 1, except for the arbitrary diagonal matrix d in the conjugation formula. But of course that formula is still true if you restrict d to be diagonal with determinant 1, so nothing really changes there.

The root system Φ also does not change. In the general theory, this comes from the fact that we have a short exact sequence

$$1 \to \mathrm{SL}_n(\mathbb{R}) \hookrightarrow \mathrm{GL}_n(\mathbb{R}) \xrightarrow{\det} \mathbb{R}^\times \to 1$$

where the left map is the inclusion, and the right map is the determinant. From this sequence, we can see that $\mathrm{SL}_n(\mathbb{R})$ is (isomorphic to) the quotient $\mathrm{GL}_n(\mathbb{R})/\mathbb{R}^\times$, and from this and some general theory it follows that the root systems of $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{SL}_n(\mathbb{R})$ have to be the same.

More generally, everything done here for $\mathrm{GL}_n(\mathbb{R})$ can be done for $\mathrm{GL}_n(R)$ (or $\mathrm{SL}_n(R)$) where R is any commutative ring with unity. If this seems excessively abstract, this means it can be done for $\mathrm{GL}_n(\mathbb{Z})$. The replacement for the nonzero real numbers $\mathbb{R}^\times$ is the group of units of R, not the set of nonzero elements of R. For example, if $R = \mathbb{Z}$, the set of units is $\{\pm 1\}$, so the diagonal subgroup of $\mathrm{GL}_n(\mathbb{Z})$ consists of matrices with only ± 1 along the diagonal. For $\mathrm{GL}_n(R)$, the diagonal subgroup, character functions, and elementary matrices are all defined as above, and the exponential, conjugation, and commutator formulas still hold. That is to say, when dealing with a ring R, the all the coefficients necessary to express these formulas come from the original ring R.

The full scope of the butterfly analogy extends to a class of objects called "split reductive algebraic groups." Given such a group G, it has a maximal torus (analog of D), Borel subgroups (analogs of U and L), characters of the maximal torus, pinning functions, and a root system. There are analogs of the exponential formula expressing something about pinning maps, the conjugation formula expressing a relationship between conjugation of pinning functions by the maximal torus and roots, and a Chevalley commutator formula expressing relationships between pinning functions controlled by the root system. Here is a table relating some of the parts of $\mathrm{GL}_n(\mathbb{R})$ to more general terminology used in books and papers.

General theory	Specific case $\mathrm{GL}_n(\mathbb{R})$
Split reductive algebraic group	$\mathrm{GL}_n(\mathbb{R})$
Maximal torus	Diagonal subgroup D
Borel subgroups	Upper/lower triangular subgroups U and L
Characters of the maximal torus	Character functions α_i
Pinning functions/root subgroup maps	Pinning functions e_{ij}
Root system	Root system Φ