One of the biggest reasons I knew I should not pursue a musical career is my historical inability to understand *counterpoint*, "the relationship between two or more musical lines which are harmonically interdependent[1]," and the techniques of *voice leading*, "the linear progression of individual melodic lines and their interaction with one another to create harmonies[2]." This procedure has always been explained to me graphically and logically, with a lot of rules and laws which provided shortcuts on how to move from chord to chord, as well as forbidden actions which you shall never do. For example, using *parallel fifths* while voice leading is very taboo; and Bach had good reason for not using them: it indeed sounds nasty, and the whole point of creating this rules for voice leading is too make sure we can quickly create movements in music which sounded nice together. Thankfully, we now have some tools to understand voice leading mathematically. In his article, *The Geometry of Musical Chords*, Dmitri Tymoczko lays down a foundation for analyzing voice leading through the lenses of topology and geometry. We begin with a few definitions and constructions.

Firstly, we can model pitches with real numbers. Sound, in real life, is described with the frequency of the vibrations of sound waves on a continuous spectrum. Therefore, we can find continuous maps which takes the frequency spectrum and assigns real numbers to them as desired. If we identify the Western 12 tones[3] to the first 12 nonnegative integers like so:

$$\begin{aligned} C &:= 0 \\ C^\sharp = D^\flat &:= 1 \\ &\vdots \\ A^\sharp = B^\flat &:= t \text{ (for ten)} \\ B &:= e \text{ (eleven)} \end{aligned}$$

We can observe a familiar equivalence relation, $\mathbb{R}/12\mathbb{Z}$, with these identifications. An equivalence class $[a]$, which we call *pitch classes*, are sets of real numbers $\{a + 12b \mid b \in \mathbb{Z}\}$ where a is some pitch. In other words, if we have a pitch $G = 7$, then the note two octaves above, $7 + 12(2) = 31 \equiv 7 \mod 12$, is still a G. Therefore, $\mathbb{R}/12\mathbb{Z}$ as the space of pitch classes is indeed a quotient space obtained from the general pitch space $\mathbb{R}$. One such map is given by Tymoczko:

$$p = 12 \log_2(f/440) + 69$$

where p is the real number assigned to the sound with fundamental frequency f (p.72, Tymoczko, 2006). This maps the conventional Middle C, at 256 Hertz, to the real number 60, which in our congruence class is equivalent to 0 mod 12. One last remark to make is that this space is still continuous, and not discrete like the 12 names of pitches. Sounds which have frequencies that are assigned to real numbers which are not integers are still valid sounds, but are not given names in a universal and convenient way in the most conventional Western studies of music. The Arab Tone System, for example, has 24 tones named as opposed to 12. The choice of having 12 tones, though not entirely arbitrary, is not a requirement but merely a choice for the convenience of this studies.

[1] https://en.wikipedia.org/wiki/Counterpoint
[2] https://en.wikipedia.org/wiki/Voice_leading
[3] The 12 tones are represented with the first 7 letters of the English alphabet with some modifications for half-steps using the sharp symbol ♯ and the flat symbol ♭. Specifically, they are: C, C♯/D♭, D, D♯/E♭, E, F, F♯/G♭, G, G♯/A♭, A, A♯/B♭, B.

With names for our pitches and a nice space to operate over, we can begin to define a voice leading and a chord. We feel naturally inclined to define a *chord* as an ordered list of pitches in the pitch class space, and a *voice leading* a multiset of ordered pairs of pitches. For example, a voice leading between the C Major chord containing the three notes C, E, and G, which can be written now as $(0, 4, 7)$, and the C Minor chord, $(0, 3, 7)$, is the multiset $\{(0, 0), (4, 3), (7, 7)\}$, commonly denoted as $(0, 4, 7) \to (0, 3, 7)$. We observe that C Major exists in the space $(\mathbb{R}/12\mathbb{Z})^3$, which is a Cartesian product of quotient spaces. However, it is not the quotient space we want. Notice that in $(\mathbb{R}/12\mathbb{Z})^3$, elements are ordered lists. This means that despite having the same three notes, $(0, 4, 7)$ is a different chord from $(0, 7, 4)$. Both are conventionally known as the C Major chord (they have the same three notes), and will sound very similarly to a human ear. We say that these two chords, which are permutations of each other with the same notes, have the same *quality* in the same key. While practicing voice leading, we are often given a list of chords, described only with their key and quality, and expected to place notes of the chords in an order such that transitions between chords are smooth and pleasant, which happens when each individual voicing is independent and melodic. In fact, unless the bass note is given, voice leading techniques completely do not distinguish between different orderings or permutations of notes in a chord as long as they are of the same quality - the entire process of voice leading is deciding where each note goes and what order the notes of a chord comes in! Additionally, the bass note is also often only given as a guideline, and could be completely unprovided at all[4]. Therefore, the mathematical difference between these two permuted lists in this product space does not accurately reflect real-life phenomenon. To avoid such complications and to consider voice leading in a broader and more accurate sense, we want to treat chords as *un*ordered multisets of pitches, where each element of $(\mathbb{R}/12\mathbb{Z})^3$ are identified with all permutations of its components: let σ be any permutation on 3 elements, i.e. a bijection from $\{1, 2, 3\}$ to itself, then for any point $a = (a_1, a_2, a_3) \in (\mathbb{R}/12\mathbb{Z})^3$, we identify a with $a_\sigma = (a_{\sigma(1)}, a_{\sigma(2)}, a_{\sigma(3)})$, for all possible σ. This identification, in fact, creates new symmetries and a new quotient space, namely $\mathbb{T}^3/\mathcal{S}_3$, the 3-torus $\mathbb{T}^3$ modulo $\mathcal{S}_3$, the symmetric group. This space is more than a simple quotient space; it is an orbifold, "a topological space locally modeled on quotients of $\mathbb{R}^n$ by a finite group action" (Caramello Jr., 2019). A similar construction holds for chords with n notes in the space $(\mathbb{R}/12\mathbb{Z})^n$, for all $n \in \mathbb{Z}_+$.

Picture! For a more concrete visualization, lets consider the 2-dimensional example given by Tymoczko. Figure 4 shows the orbifold $\mathbb{T}^2/\mathcal{S}_2$, which initially derives from $\mathbb{R}^2$; let $\mathbb{Y} = \mathbb{T}^2/\mathcal{S}_2$ for convenience. First visualize, in the first quadrant of $\mathbb{R}^2$, a few points[5] as shown below. Applying the quotient map which glues together points that differ by exactly 12 in either component, we give it the equivalence class of congruence modulo 12: in Figure 1, points which are identified together are shown in the same colours. Specifically, any $(a, b) \equiv (3, 6)$ are labeled red, any $(a, b) \equiv (6, 3)$ are labeled orange, and any $(a, b) \equiv (9, 9)$ are labeled yellow. Recalling our definition of chords, we also want to apply a new equivalence relation on this: $(a, b) \sim (c, d)$ if $(c, d) = (a, b)$ or $(c, d) = (b, a)$. Here is a quick proof that this indeed defines an equivalence relation.

[4]This is a whole other technical discussion of the role of the bass note, and what a bass note even is. However, the choice of bass note is often decided by general voice leading techniques, instead of by conscious or arbitrary choice. Therefore, it is well justified to disregard this slight detail about the choice of bass note and go forward in the most general sense, because it is sufficient and not unrealistic to use broader definitions and techniques on bass notes as well.

[5]Namely $(3, 6)$, $(6, 3)$, $(9, 9)$, $(15, 6)$, $(18, 3)$, $(21, 9)$, $(27, 6)$, and $(30, 3)$.

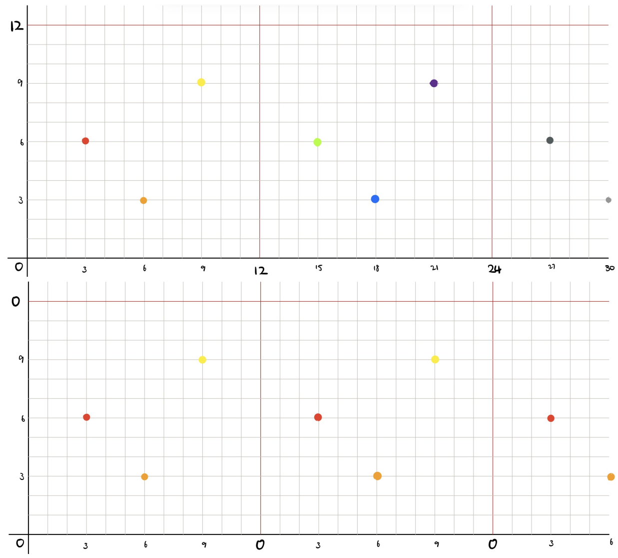

Figure 1: $\mathbb{R}^2$ and $\mathbb{R}/12\mathbb{Z}$ with the same 8 points.

Proof:

1. Reflexivity. Consider (a,b) and compare it to some $(c,d) = (a,b)$. Then $(a,b) \sim (c,d)$, so $(a,b) \sim (a,b)$.

2. Symmetry. Suppose $(a,b) \sim (c,d)$. Then either $c = a, d = b$ or $c = b, d = a$. Then, we know $(c,d) \sim (a,b)$, since either $a = c, b = d$ or $a = d, b = c$.

3. Transitivity. Suppose $(a,b) \sim (c,d)$ and $(c,d) \sim (e,f)$. Then, if $(c,d) = (a,b)$, then either $(e,f) = (c,d) = (a,b)$ or $(e,f) = (d,c) = (b,a)$, which implies that $(a,b) \sim (e,f)$. Else, if $(c,d) = (b,a)$, then either $(e,f) = (d,c) = (a,b)$ or $(e,f) = (c,d) = (b,a)$, which also implies that $(a,b) \sim (e,f)$.

$\blacksquare$

As the bigger picture offers to us no more information than the 12×12 square, which is the quotient

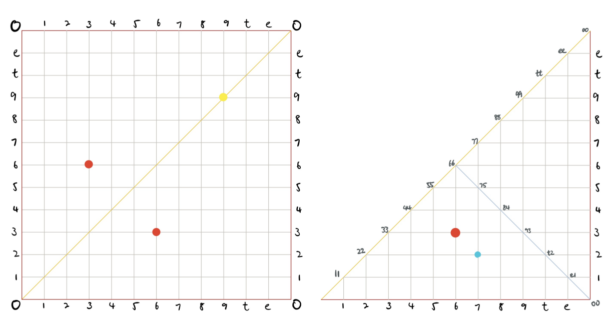

Figure 2: New gluing on the left, new quotient space on the right.

space, outlined by thin red lines, we can simply investigate the square for simplicity. Notice that our new equivalence relation does not glue points on the line $y = x$ to any other points, while points on either side of $y = x$ will be glued to the point which is its reflection from the line. Identifications, then, are made as in Figure 2: $(3,6) \sim (6,3)$, shown in red, while $(9,9)$, shown in yellow, does not get glued anywhere else.Once again we can simplify this space by only considering the new quotient

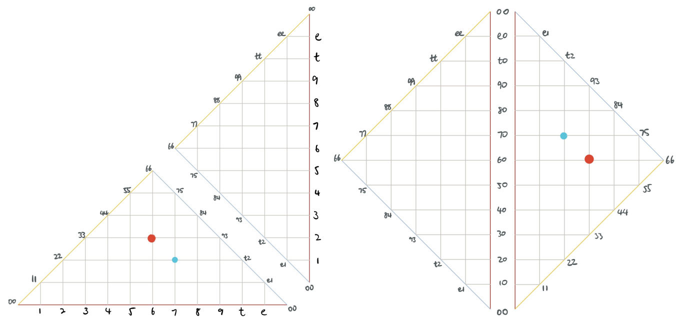

Figure 3: First separate into two triangles, then align red edges.

space, which is one of the two triangles of this red square, as shown in Figure 2. Noticing that the vertical red line on the right and the horizontal red line on the bottom are identical, we cut up this triangle into two smaller isosceles triangles along the middle line segment which connects $\{6,6\}$, denoted 66, to $\{0,0\}$, denoted 00, through points 75, 84, 93, $t2$, and $e1$. Now we rotate the bottom triangle clockwise by 90 degrees, and "flip" it by 180 degrees. We can visualize this flip with the help of marking an extra point, 72, in blue, and tracking where it lands in the final alignment in Figure 3. Finally, we observe, on the left side of Figure 4, that the two opposite yellow edges, which go from 00 through 33 to 66 and from 66 through 99 to 00, are different. However, the two blue edges, which both go from 00 through 93 to 66, are identical. The tricky part, however, is that they are oriented with half-twist's difference. This translates to the familiar square as shown on the right side of Figure 4.

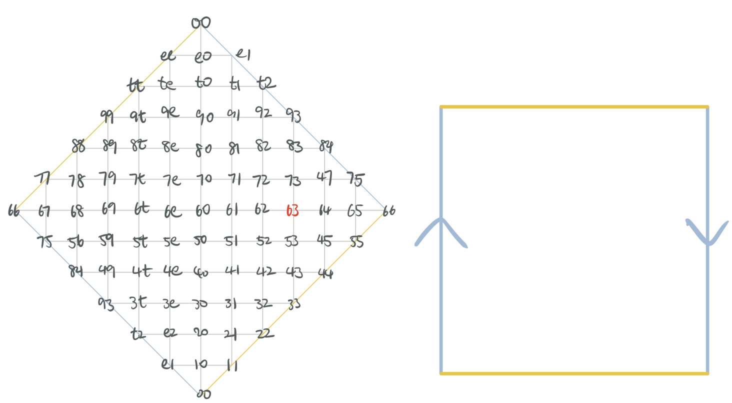

Figure 4: Each point labeled in the new square, $\mathbb{Y}$; identification of sides of new square.

Generally, we want independent and bijective voice leadings. A voice leading is *independent* if it is not a transposition, i.e. if it does not look like $(a_1, ..., a_n) \to (a_1 + b, ..., a_n + b)$ for some fixed value b. We will see the reasoning in a short moment. A voice leading being *bijective* means exactly what we expect it to mean: injectivity and surjectivity guarantees that each note in the first chord moves uniquely to a note in the second chord, and each note in the second chord has a fiber of one element from the first chord. When we use the word "voice-leading," we usually have a fixed number of voices. For example, soprano, alto, tenor, bass, is a standard four-part set-up of a choral piece of music. In this case, we are determining the movements of four-note chords, who live in $\mathbb{T}^4/\mathcal{S}_2$. By requiring that all voice leadings are bijective, we guarantee that the music will be able to keep flowing, without parts disappearing or spontaneously appearing. What we want to do with each of these four parts is to minimize individual movement. It is not only difficult, but also unpleasant, to listen to someone sing drastically different notes consecutively. Imagine someone singing “somewhere” from “somewhere over the rainbow,” and going back and forth singing only those two notes - it is simply not melodic and enjoyable. The most intricate and beautiful pieces of music will have sensible melodies in each part, while still maintaining a functional harmony on the large scale. Therefore, given a list of harmonies, we want to arrange the notes in each chord according to their parts while minimizing the distance between consecutive notes.

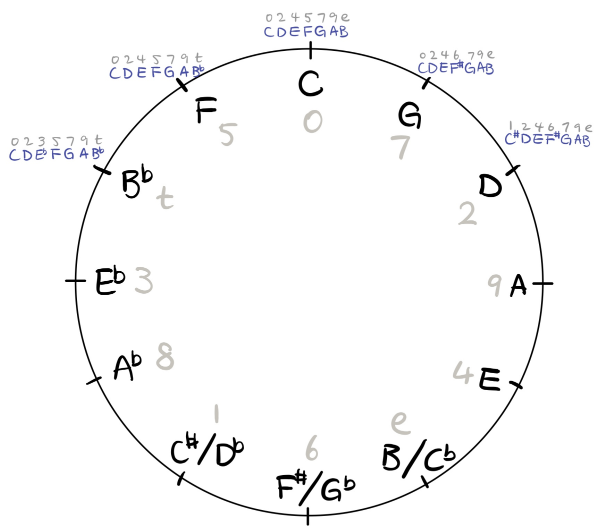

Figure 5: The Circle of Fifths.

An important tool that classical musicians use is the “Circle of Fifths,” which places the 12 tones in a circle, starting with C on top in the middle, then moving a “perfect fifth” or 7 half-steps above, we have G to the right side of C. Then, a fifth above G is D ($7 + 7 = 14 \equiv 2 \mod 12$), etc. See Figure 5. Often, at the end of a piece, we want to arrive at the tonic chord, or “home” or main chord of a piece, in a satisfying way. If the piece of music is in C major, a common way to end the piece is the progression pattern of $A - D - G - C$, which we observe is just going along the Circle of Fifths consecutively until we reach C at the top. Such an ending, a *cadence*, is often called a *perfect cadence*. What makes it perfect, other than the fact that it sounds nice? The answer is that a perfect cadence such as $D - G - C$ in C major uses very efficient voice leadings which are easily independent and bijective. Notice that to go from the G Major scale $\mathbf{G} = \{G, A, B, C, D, E, F^\sharp\}$ to the C Major scale $\mathbf{C} = \{C, D, E, F, G, A, B\}$, we only need to to flatten the $F^\sharp$ to an F in order to make the two multisets equal. This means that we can define a voice leading which is almost the identity, on the pitch space restricted to $\mathbf{G}$, except at $F^\sharp$, which in turn guarantees that the voice leading is independent. In terms of music theory, moving from the chord based at the fifth scale degree of the main tonic chord (i.e. what G is to C) is desirable precisely because of how close the scale is to the tonic scale: it creates a slight tension such that when we do transition into the tonic chord, the release of the tension feels satisfying. Therefore, the Circle of Fifths works, with the smallest unit of change between each step of the circle, to fully move the music along in a logical and pleasant way.

Now let us direct our attention back to why parallel fifths are generally considered bad voice leading. Consider $\mathbb{Y}$, as visualized above. Given that the tonic or root of the scale and the fifth of the scale are the most important, consider the progression $(29, 72, 07)$, which is not only is a $D - G - C$ progression that follows the Circle of Fifths, but also are three chords which are a "fifth" apart distance-wise. If we choose to voice lead these chords in a way such that motions are parallel and not independent, we would get the voice leadings $(2, 9) \to (7, 2) \to (0, 7)$, which is a lot of movement. We move by 5 units each step in each component, which means that we are travelling a lot very unnecessarily. Inspecting Figure 6, we see that if we choose parallel motion for voice leading

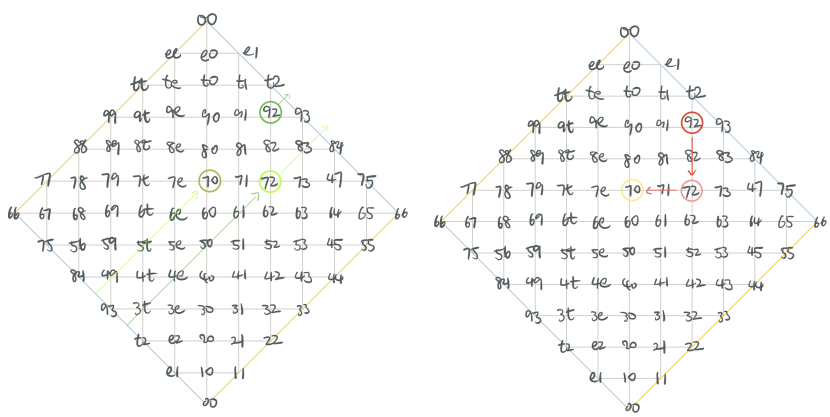

Figure 6: Two different voice leadings of the same progression.

around the Circle of Fifths, we will be quickly and exhaustively travelling around the Mobius Band: in just 3 chords, we almost make it an entire time around the surface of the band. However, if we chose to voice lead while minimizing movement, then we can save the trouble by keeping the repeated note in place, and only moving the different note. Compare the parallel movements, in green, to the minimized movements, in red.

This was a perfect example of how math shows up in the least expected places.

References.

Caramello Jr., Francisco C. “Introduction to Orbifolds.” 10-12.

Tymoczko, Dmitri. “The Geometry of Musical Chords.” *Science* 313: 72-74.