A Geometric Construction for the Thornton Primaries

Paul Centore

Abstract

Colour-matching functions (CMFs) in standard colour-matching experiments consistently peak at three wavelengths, known as the Thornton primaries. This paper derives the CMFs' peaks from a geometric construction, the unit-power hull U, whose shape explains the observed consistency. The spectrum locus vectors consist of all the colours (as vectors in three-dimensional colour space) that arise from single-wavelength spectral power distributions (SPDs) of some unit power. The convex hull of the spectrum locus vectors will be called the unit-power hull U; it is the set of all colours that can be produced with unit power or less. Now choose three monochromatic SPDs (reasonably spread out over the spectrum) for a colour-matching experiment; equivalently, assign a basis to three-dimensional colour space. To find the peak for one primary's CMF, use the vectors for the other two primaries to generate a plane in colour space. Parallel-translate that plane across U; the farthest possible translation point is the tip of the spectrum locus vector for the peak wavelength of the first primary's CMF. The relative pointedness of U at these vertices limits the set of farthest points to three narrow wavelength bands, centered on the Thornton primaries.

Index Terms: Thornton primaries, primary, colour-matching function, spectrum locus, convex hull

1 Introduction

The human retina's long-, medium-, and short-wavelength cones respond with varying intensity to different wavelengths in the visible spectrum (between about 400 and 700 nm). A cone response curve gives the intensity of a cone's output as a function $f(\lambda)$, where λ denotes wavelength. Since no one has yet measured cone responses directly, a mathematically equivalent colour-matching experiment is used, in which observers view a bipartite field, consisting of two adjacent spectral power distributions (SPDs), each of which has a certain physical power level as a function $s(\lambda)$ of wavelength λ. Even when the SPDs are very different physically, a human observer might perceive their colours as identical.

In a typical colour-matching experiment, one half of the field displays a monochromatic SPD, i.e. one whose power is restricted to a single wavelength. In the other half, an observer superposes three other monochromatic SPDs, called primaries, varying their power levels until a colour match is obtained. (In case no match can be made, one of the three primaries

can be transferred to the first half of the field, and power levels adjusted until a match occurs; the transferred primary is then assigned a negative power level.) The observer steps through individual target wavelengths, typically from 400 to 700 nm in steps of 10 nm, making a match at each step. A colour-matching function (CMF) for a certain primary specifies, for each wavelength, the power level required of that primary in the colour match. A bipartite colour-matching experiment therefore results in three CMFs, all of which are functions over the visible spectrum. Further mathematical analysis allows each colour to be expressed as a vector in a three-dimensional vector space V, sometimes called *colour space*. Not every vector in V corresponds to a real colour; some vectors are interpreted as imaginary colours, that cannot be produced physically.

Many different sets of primaries can be used, each producing a different set of CMFs. In fact, though, the CMFs for one set of primaries can be mathematically calculated from a second set (Sect. 4.4 of Ref. 1). Furthermore, one can even work with imaginary primaries, whose SPDs require negative power levels at some wavelengths. Though experiments with imaginary primaries cannot actually be performed, their CMFs can be calculated from the CMFs for real primaries. A well-known example is the CMFs $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$, standardized in 1931 by the Commission Internationale de l'Éclairage (CIE),[2] and shown in Figure 1.

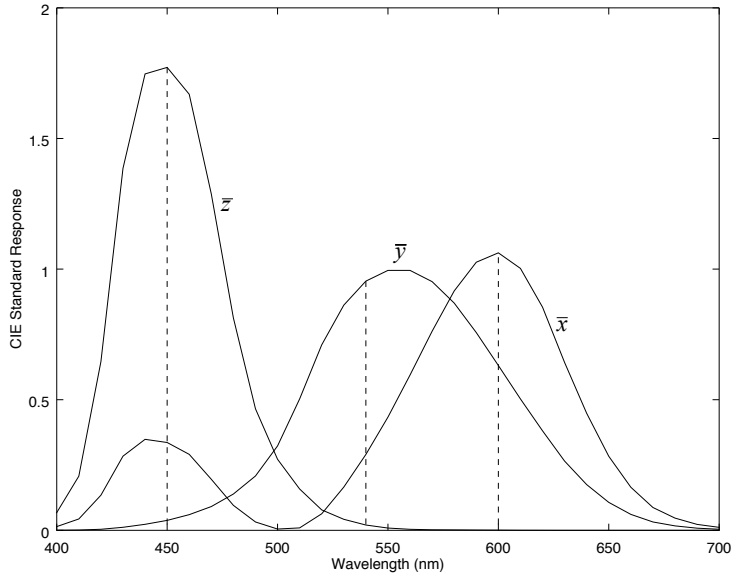

Figure 1: The CIE 1931 Colour-Matching Functions

Like others before him, Thornton[3] examined plots of many CMFs, produced from many different sets of primaries. Each CMF seemed to have a clear peak at a single wavelength, so any set of primaries produced three distinct peak wavelengths. Furthermore, the peaks varied only slightly when the the primary stimuli were changed. By averaging the peak wavelengths of a large set of CMFs, Thornton arrived at three wavelengths, now called the Thornton primaries, which we will denote by T_1, T_2, and T_3. If the observed field subtends 2 degrees as it does for the 1931 standard, then the Thornton primaries are 447, 541, and

604 nm; for a field that subtends 10 degrees, they are 446, 538, and 600 nm (see Table 1 of Ref. 3).

This consistency of peak wavelengths raises some questions. To begin with, the 447 nm primary seems more stable than the other two: Thornton (again, see Table 1 of Ref. 3) found an average deviation of only 0.4 nm around 447 nm, but average deviations of 3 or 4 nm around 541 and 604 nm. Why should there be such a difference? Furthermore, the Thornton primaries seem to occur strongly for monochromatic primaries, but more weakly for other primaries, especially imaginary ones. Even though the CMFs $\bar{x}(\lambda)$ and $\bar{z}(\lambda)$ in Figure 1 peak very near the Thornton primaries, for instance, the CMF $\bar{y}(\lambda)$ peaks at 555 nm, a full 14 nm away from the 541 nm Thornton primary. Why do Thornton's results hold more or less strongly for different primaries chosen for a colour-matching experiment?

The three-dimensional shape of the geometric construction presented in this paper answers these questions, showing visually why the Thornton primaries appear regularly, but not invariably. When working in V, the colours corresponding to the Thornton primaries form a figure that we will call the Thornton tetrahedron, $\mathcal{T}$. Another important object in colour space is the unit-power hull $\mathcal{U}$, which consists of all the colours that can be produced by an SPD whose total power is no more than some unit power. $\mathcal{U}$ is a convex polyhedron, each of whose vertices is the colour vector for a unit-power monochromatic SPD. Figure 2 shows $\mathcal{U}$, in multiple views, and Figure 3 shows $\mathcal{T}$ inside $\mathcal{U}$, where it can be seen to be a good approximation.

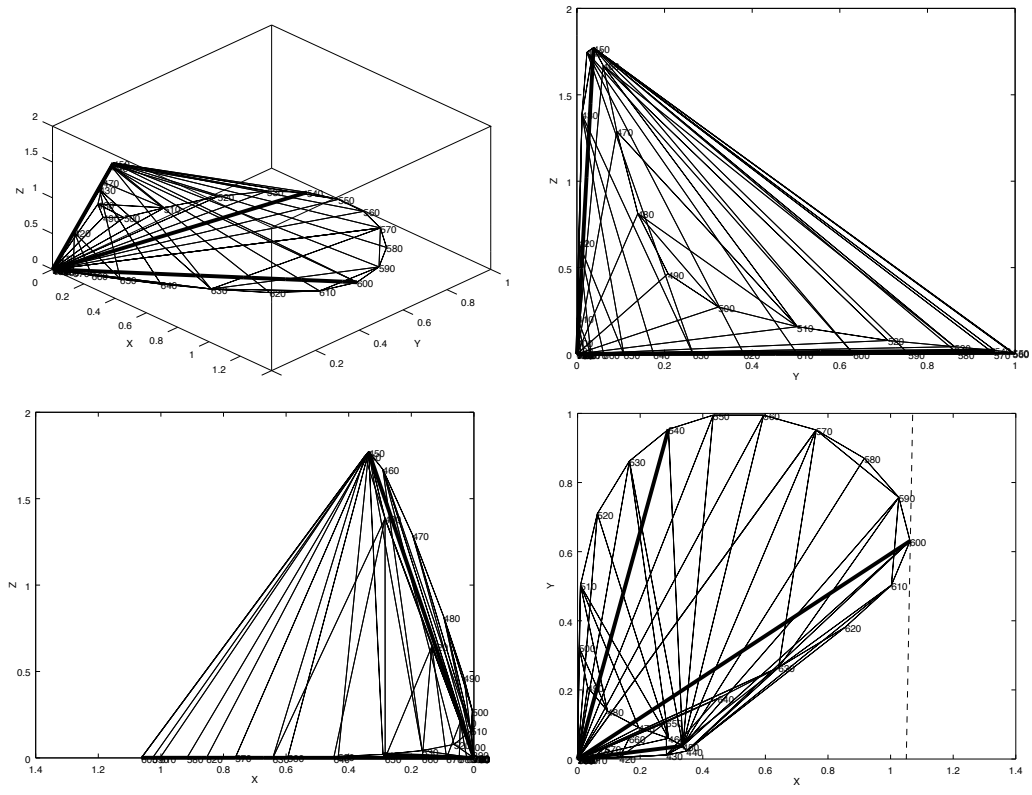

Figure 2: The Unit-Power Hull $\mathcal{U}$ (Top Right: View along X-axis; Bottom Left: View along Y-axis; Bottom Right: View along Z-axis)

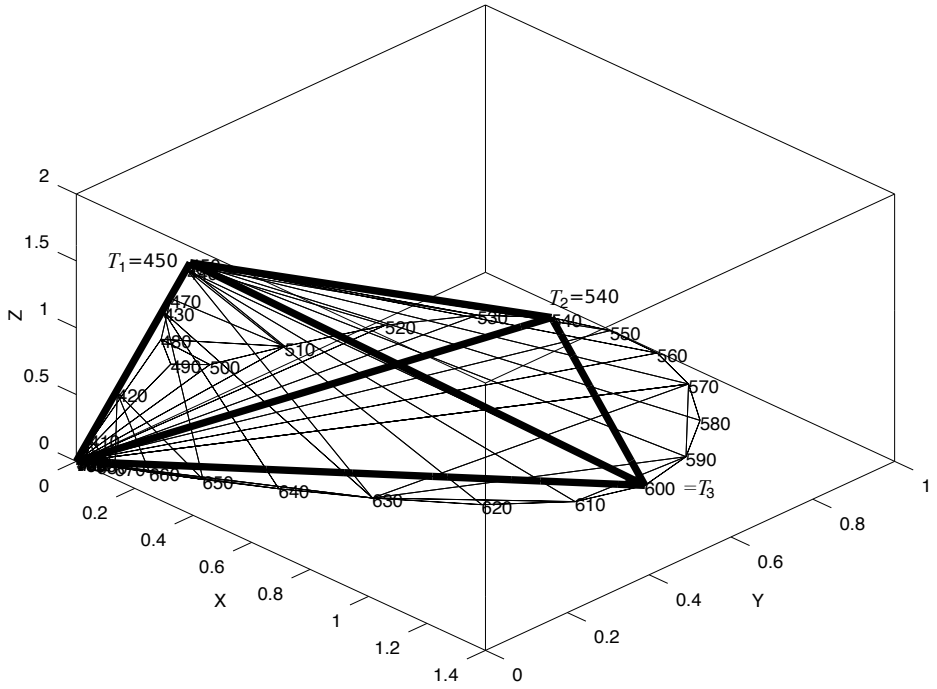

Figure 3: The Thornton Tetrahedron $\mathcal{T}$ as an Approximation to the Unit-Power Hull $\mathcal{U}$

This paper will demonstrate mathematically that the peak wavelengths can be expressed as solutions to a linear programming (LP) problem on $\mathcal{U}$. Standard linear programming theory (Theorem 1.7 of Ref. 4) tells us that LP solutions typically occur on the vertices, which in this case are associated with wavelengths. Every set of primaries leads to its own LP problem: each pair of primaries defines a plane, which we aim to translate as far as possible across $\mathcal{U}$. The farthest stopping point occurs at a vertex, and the wavelength associated with that vertex is the peak wavelength for the primary that was *not* used in constructing the plane.

One noticeably pointed vertex occurs at 450 nm, which is the first Thornton primary (rounded to the nearest 10 nm). The pointedness of the 450 nm vertex allows it to be the solution to many LP problems, associated with many sets of primaries, which explains why the first Thornton primary varies so little. The vertices for the other two Thornton primaries are more rounded, leading to more variation in the LP solutions, and thus more variation in the CMF peaks.

The three primaries for a colour-matching experiment can be drawn as basis vectors in the three-dimensional colour space V, where $\mathcal{T}$ and $\mathcal{U}$ both live. Experimenters usually choose primaries that are monochromatic and widely spaced, which makes them fairly close to $\mathcal{T}$—at least in the sense that their LP problems have very similar solutions. This similarity explains why CMFs so regularly peak close to the Thornton primaries. The CMF peaks for a closely spaced set of primaries, such as 490, 500, and 510 nm, would deviate significantly from Thornton's wavelengths, but experimenters would likely not choose such a set. Imaginary

primaries, of course, could never be used, but if one calculates CMFs for imaginary primaries mathematically, then, like the CIE primaries, their peaks can also deviate significantly. Geometrically, the vector for an imaginary primary could be very far outside $\mathcal{U}$, whereas the vectors for real primaries are always contained in $\mathcal{U}$. Thus two facts—that monochromatic, widely spaced primaries tend to occur in practice, and that $\mathcal{T}$ approximates $\mathcal{U}$ well—explain why Thornton's wavelengths occur so frequently.

The paper is organized as follows. First, the relevant colour science objects, such as $\mathcal{U}$, are constructed from first principles. Second, since $\mathcal{U}$ lives naturally in a three-dimensional vector space, and some further derivations rely solely on linear algebra rather than colour science *per se*, some geometric constructions are described that require only a vector space setting. Third, the vector space constructions are applied to peak wavelengths; for ease of understanding, important geometric concepts are illustrated with two-dimensional examples. Fourth, the foregoing development is applied directly to the Thornton primaries. Fifth, a geometric analysis and discussion leads to this introduction's explanations for the regular appearance of Thornton's primaries. Finally, some previous results are derived using the new geometric construction, and a summary is given.

2 Color Science Constructions

The human vision system converts physical SPDs to colour perceptions. An SPD can be expressed as a function $s(\lambda)$, in some units of power, over the wavelengths λ in the visible spectrum, from about 400 to 700 nm. Following common practice, an SPD function in this paper will be expressed as a discrete set of 31 values at increments of 10 nm from 400 to 700 nm. SPDs can be classified as *monochromatic* if all their power is restricted to a single wavelength (or in practice, a narrow band of wavelengths), and *polychromatic* otherwise. The monochromatic SPD of unit power at wavelength λ will be denoted 1_λ. SPDs can be added as functions, or multiplied by a scalar, producing in either case new SPDs; if one allows the mathematical fiction of negative power levels, then the set S of SPDs takes on the linear structure of a vector space. The 31 monochromatic SPDs define a convenient basis for S.

In a classical colour-matching experiment, two different SPDs are displayed side-by-side, and an observer adjusts them until their colours appear identical. Notably, very different SPDs can produce exactly the same colour. A *colour* in fact, can be defined as an equivalence class of SPDs that produce an identical perception, so any SPD can be assigned to a colour unambiguously. Furthermore, Grassmann's laws (Sect. 4.3.2 of Ref. 1) imply that this assignment is a linear transformation, so the set C of all colours must be a subset of a vector space V, and empirical research has determined that V, which is sometimes called *colour space*, is three-dimensional. Using this transformation, any SPD can be assigned to a three-dimensional colour vector in V.

In 1931, the Commission Internationale de l'Éclairage (CIE) summarized the results of bipartite experiments in their Standard Observer.[2] A standard set of coordinates, denoted X, Y, and Z, was chosen for the vector space V. Given any SPD $s(\lambda)$, the CIE coordinates

for the colour vector of $s(\lambda)$ are given by

$$X(s) = \sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{x}(\lambda), \tag{1}$$

$$Y(s) = \sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{y}(\lambda), \tag{2}$$

$$Z(s) = \sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{z}(\lambda), \tag{3}$$

where $\bar{x}$, $\bar{y}$, and $\bar{z}$ are colour-matching functions (CMFs) that the CIE also standardized.

Monochromatic SPDs bear a simple yet important relationship to CIE coordinates:

$$X(1_\lambda) = \bar{x}(\lambda), \tag{4}$$

$$Y(1_\lambda) = \bar{y}(\lambda), \tag{5}$$

$$Z(1_\lambda) = \bar{z}(\lambda), \tag{6}$$

because 1_λ takes on the value 1 at wavelength λ and 0 everywhere else. The foregoing equations show that X, Y, and Z can be interpreted not just as coordinates, but also as linear functionals, on either the vector space S of SPDs, or colour space V.

Mathematically, fixing coordinates is equivalent to choosing three basis vectors; in our context, a vector is a colour, and a basis vector is called a *primary colour* or just a *primary*. The primary P_X with coordinates $XYZ = (1, 0, 0)$ consists of a set of perceptually identical SPDs, all of which have negative power at some wavelengths, so it is called an imaginary colour—it is a mathematical solution that does not exist in the real world. The primary P_Y that produces a positive Y but zero X and Z is similarly imaginary, as is the primary P_Z that produces a positive Z but no X or Y. Despite the negative power levels, however, the associated CMFs (i.e. $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$) are non-negative at every wavelength.

Any set of three linearly independent primary colours can constitute a basis for V. In a typical bipartite field experiment, a linear combination of the primaries fills one side of the field, and a target colour fills the other side. A subject adjusts the coefficients in the linear combination until the two sides appear identical. Usually, the target colours cycle through 1_{400}, 1_{410}, etc., until 1_{700}, in which case the coefficients make up the CMFs for that set of primaries. If a colour match is impossible, then one of the primaries is added to the target colour, and a match is made with some combination of the remaining two primaries; the coefficient of the transferred primary is then set to a negative.

Many, in fact most, of the vectors in V are imaginary colours that exist mathematically but cannot be produced by any SPD whose power level is non-negative at every wavelength. Since each SPD is a positive linear combination of monochromatic SPDs, and since S is linearly embedded in V, the colours in V that can be physically produced are non-negative linear combinations of the colours resulting from monochromatic SPDs, all of which have CIE coordinates

$$(X(1_\lambda), Y(1_\lambda), Z(1_\lambda)) = (\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)) \tag{7}$$

for some wavelength λ. Such vectors are called *spectrum locus vectors*, and the curve their tips trace out is called the *spectrum locus*. As one moves to the ends (or slightly beyond the ends) of the visible spectrum, all the cone responses decrease to zero, so the spectrum locus vectors approach the zero vector arbitrarily closely; the zero vector will therefore also be considered a spectrum locus vector. The set of all non-negative linear combinations, also called the *convex cone*, of the spectrum locus vectors is just the set C of all physically possible colours.

A more important set for us is the set $\mathcal{U}$ of all colours produced with no more than unit power. Algebraically, $\mathcal{U}$ consists of all linear combinations of the spectrum locus vectors whose coefficients are non-negative and sum to 1. Geometrically, $\mathcal{U}$ is the convex hull of the spectrum locus vectors. Furthermore, standard convexity theory (Sect. 20 of Ref. 5) tells us that $\mathcal{U}$ is a polyhedron, every vertex of which is the tip of the spectrum locus vector for some wavelength λ. Figure 2 shows the unit-power hull. The upper left corner shows an off-center view while the remaining three corners show the three elevations which occur when looking along the X-, Y-, or Z-axes. In the upper right corner, for example, the X-axis disappears as it is pointing directly at the viewer's eye, but the Y-Z-plane is fully visible, and one sees the projection of $\mathcal{U}$ onto that plane.

This section has converted colour-matching data into some objects of colour science in a three-dimensional vector space. The next section will develop some mathematical tools that depend solely on a vector space setting, and further sections will apply those tools to the new objects to reach conclusions about the Thornton primaries.

3 The Vector Space Setting

3.1 Geometric Constructions

The geometric insights that this paper relies on are naturally expressed in a vector space setting. This section illustrates them with easily understood two-dimensional examples, which will then be generalized to three dimensions and applied to the relevant colour science objects.

Consider a convex polygon $\mathcal{P}$ in a two-dimensional vector space coordinatized by x- and y-axes, as shown in Figure 4. The x-value of any point p of $\mathcal{P}$ can be found geometrically by parallel-translating a copy of the y-axis until it contains p, as shown in the figure. That translated copy intersects the x-axis at some value x_p, which is the x-coordinate of p.

Now suppose that we wish to find the point of $\mathcal{P}$ with the maximum x-value. Then we could translate a parallel copy of the y-axis as far as possible to the right, so that it just touches $\mathcal{P}$ at some point q; use L to denote the translate of the y-axis through q. Finding the coordinate x is equivalent to evaluating the linear functional $f(x, y) = x$, so, in terms of linear programming (LP), we are maximizing the linear functional f over a convex polygon. Standard LP theory (Sect. 1.4 of Ref. 4) tells us that the maximum typically occurs at a vertex of $\mathcal{P}$, as the figure shows. Surprisingly, this maximum construction depends much more on y than it does on x. Suppose we use a different coordinate x' for our basis, but the same coordinate y. As long as x' lies on the right-hand side of y, even if it points more up or down, or its magnitude is increased or decreased, its maximum value would still occur at q.

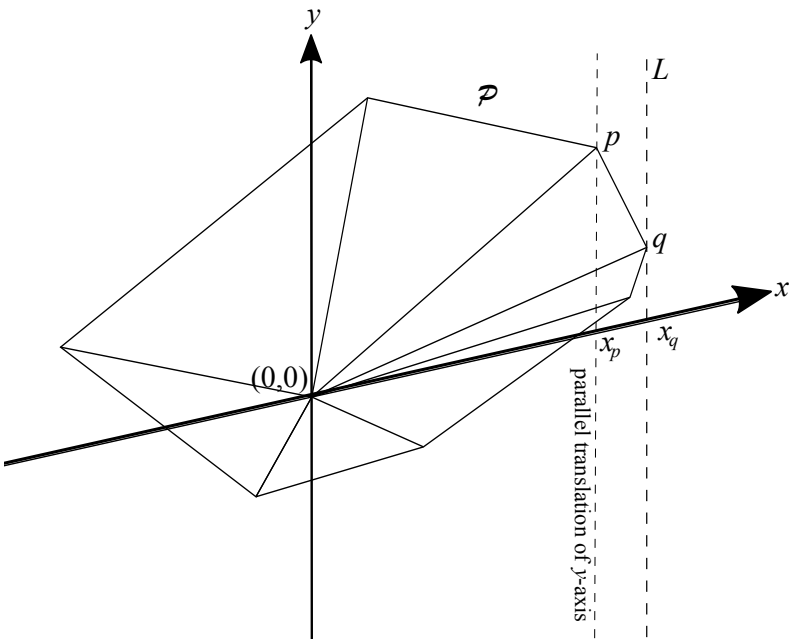

Figure 4: A Convex Polygon in the Plane

One could repeat this construction for every vertex of $\mathcal{P}$, yielding the set of parallel lines shown in Figure 5. Although there is no concept of distance between the parallel lines, distance *ratios* can still be found. The line L_2 through p_2, for example, is twice as far from the origin as the line L_1 through p_1, so the x-coordinate for p_2 is twice the x-coordinate of p_1. In fact, the vector space could be recoordinatized, leaving the y-axis fixed, but changing the x-axis to some new x'-axis, as shown in the figure. Then the x'-coordinate for p_2 would still be twice the x'-coordinate for p_1, and all other ratios between the parallel copies of the y-axis would also be preserved. This statement holds for *any* alternate basis vector, no matter in what direction it points, or how far it extends. If one plotted the x-coordinates of the vertices as a function over a line with equally spaced points indicating the vertices, then choosing a new coordinate x' would leave that function's shape unchanged, but multiply it by some scalar multiple. (This observation will be used later to prove the result[6,7] that a primary's colour-matching function maintains its shape even when that primary changes its wavelength, provided the other primaries do not change.)

The constructions in Figures 4 and 5 generalize to three dimensions, coordinatized by x, y, and z. In three dimensions, $\mathcal{P}$ becomes a convex polyhedron. Rather than parallel-translating an axis, we will parallel-translate a hyperplane, which in the case of three dimensions is just an ordinary plane, spanned by two of the basis vectors. Any point p in $\mathcal{P}$ can be projected along the parallel-translated hyperplane through p, onto a unique point on the third basis vector; this construction is equivalent to finding the third coordinate for p when p is expressed in terms of that basis. For example, the yz-plane can be translated across $\mathcal{P}$, assigning an x-coordinate to each point of $\mathcal{P}$. Just before leaving $\mathcal{P}$ the plane will intersect the point q at which x takes on its maximum value. As in the two-dimensional case, LP theory again says that q typically falls on a vertex of $\mathcal{P}$, which we can associate with the vector joining 0 to q.

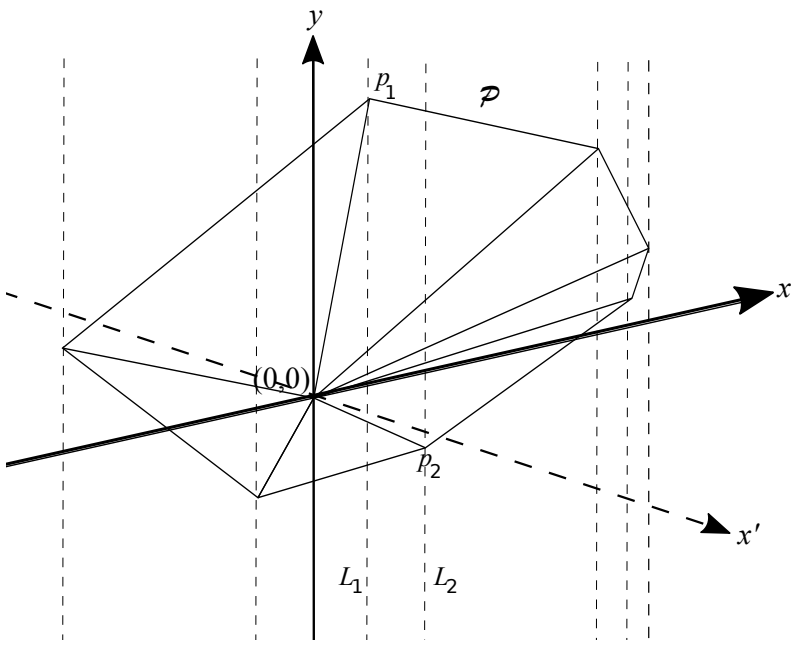

Figure 5: Relative Spacing of a Translated Axis

3.2 Application to Peak Wavelengths

The discussion in the previous section allows a geometric interpretation of peak wavelengths, which will shed some light on the Thornton primaries.

We will work in the vector space V of colours. The set $\mathcal{U}$, shown in Figure 2, of colours produced with unit power or less is the convex hull of the spectrum locus vectors, which appear as its vertices, and will be taken as $\mathcal{P}$. The CIE coordinate X (or one of the other coordinates) is a linear functional on V. To evaluate X at some colour v, draw through v a plane parallel to the Y-Z-plane. That plane will intersect the X-axis at the value $X(v)$. If v is a vertex of $\mathcal{U}$, then v is the colour associated with some 1_λ, so, by Equation (1),

$$X(v) = \bar{x}(\lambda). \tag{8}$$

To maximize X, move the parallel plane as far from the origin as possible, while still touching $\mathcal{U}$. By linear programming theory, this maximum will typically occur at some vertex of $\mathcal{U}$, which is associated with some wavelength $\lambda_{\max}$. In fact, we can see geometrically that $\bar{x}(\lambda_{\max})$ is as far from the origin as possible, so the maximum value of $\bar{x}$, i.e. its peak wavelength, occurs at $\lambda_{\max}$. We have therefore converted the problem of finding the peak wavelength into a geometric problem: translating a plane given by two basis vectors as far as possible from the origin, while still touching $\mathcal{U}$.

This geometric reformulation of peak wavelengths makes clear the dependence of CMFs and their peak wavelengths on the basis chosen for V. While $\mathcal{U}$ is a fixed subset of V, the choice of basis vectors such as X, Y, and Z is arbitrary, and one could easily select different vectors. In fact, if monochromatic primaries of wavelengths λ_1, λ_2, and λ_3, are chosen, as is often done, then the three axes will go right through $\mathcal{U}$: the axes will start at the origin, and exit $\mathcal{U}$ at the spectrum locus vectors for $1_{\lambda 1}$, $1_{\lambda 2}$, and $1_{\lambda 3}$. The planes parallel to pairs of the new axes will now also shift, as will their intersection values with the remaining axis.

The amount of the shift, though, will depend strongly not just on the choice of basis, but also on the *shape* of $\mathcal{U}$. It will be seen that a condition in which peak wavelengths do not change when the primaries change requires $\mathcal{U}$ to have a few very pointed vertices, while a smoother, more rounded shape leads to greater variation in peak wavelengths. The following examples will illustrate these statements in one or two dimensions, from which they are readily generalized to three dimensions.

3.2.1 One Response Curve

As a first example, let us suppose that a visual system only has one kind of cone instead of three, and that that cone's response curve is an inverted parabola over the visible spectrum, as seen in Figure 6. We then need only one primary, that generates one CMF. Whether that primary is monochromatic or polychromatic, the CMF will have the same shape as the response curve, up to some scalar multiple. In fact, such a one-cone system could not distinguish between wavelengths at all, but could only identify differences in brightness. (The human night-vision system actually has this structure.)

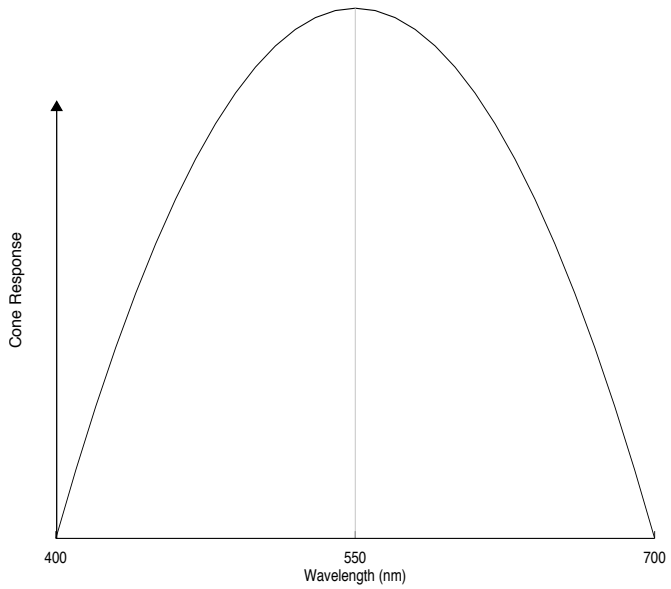

Figure 6: Response Curve for a Single Cone

Since the CMF duplicates the response curve, the peak wavelength will always occur at the wavelength that produces the greatest response (550 nm in this case), and this result holds regardless of the primary chosen. Colour space for this case would be a one-dimensional line, and $\mathcal{U}$ would be a segment of that line, from 0 to some maximum value; the maximum value is at the vertex that corresponds to the spectrum locus vector for 550 nm. All the spectrum locus vectors would lie along the same line segment, though they would differ in magnitude.

3.2.2 Two Non-Overlapping Response Curves

Let us now modify the first example by adding a second cone and supposing that the cone response curves are two inverted parabolas, one over the left half of the visible spectrum, and the other over the right, as shown in the left half of Figure 7. These curves could be called non-overlapping because any monochromatic SPD now stimulates only one of them; a polychromatic SPD, however, could stimulate both simultaneously. Let us choose as primaries two monochromatic SPDs, one in the right half and the other in the left; denote by X and Y respectively the coordinates that result when the primaries' spectrum locus vectors constitute a basis for colour space.

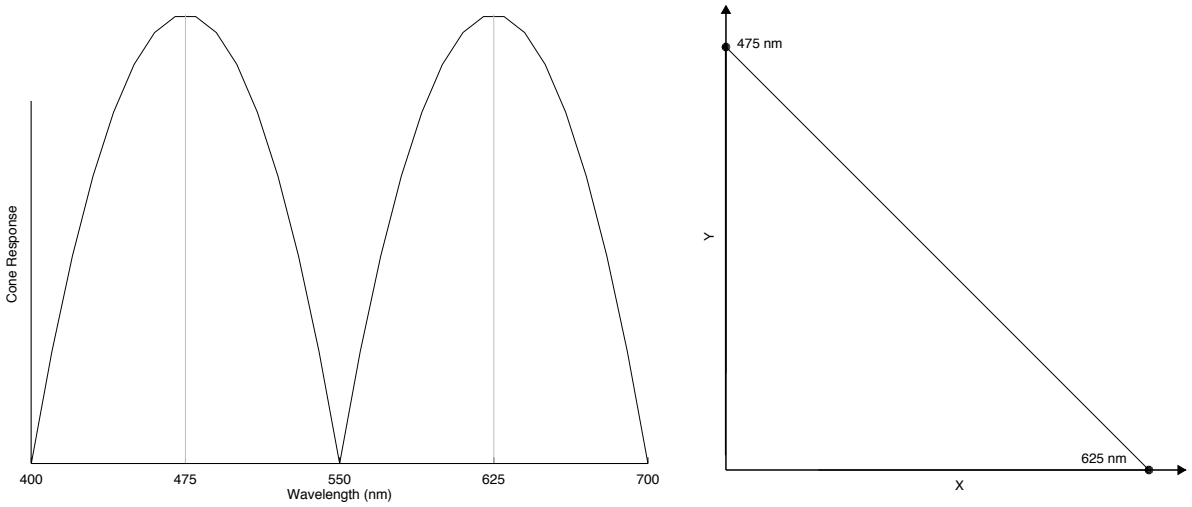

Figure 7: Response Curves and Unit-Power Colours for Two Non-Overlapping Cones

The spectrum locus vectors for wavelengths in the left half all lie along the Y-axis; they fill out a line segment whose maximum reach occurs at the wavelength 475 nm, to which the first cone responds with the most strength. A similar pattern occurs along the X-axis for the second cone. Any monochromatic SPD will stimulate only one cone, so its colour vector lies on one of the axes. A polychromatic SPD s, however, could have some power in the left half of the spectrum as well as some power in the right half, so it could stimulate both cones simultaneously. Both coordinates of its colour vector would be positive, so it would lie in the first quadrant. Since the SPDs we are considering have no more than unit power, the colour vector of s must be in the convex hull $\mathcal{U}$ of the spectrum locus vectors, all of which lie in two line segments on the X- and Y-axes. $\mathcal{U}$ is actually the triangle in the right half of Figure 7.

According to Equation (4), the value of the CMF for X at wavelength λ is just the X-coordinate of 1_λ. According to the linear algebra arguments developed earlier, the peak wavelength for X can be found by parallel-translating the Y-axis until it reaches a line L (shown in Figure 8), which is as far as possible to the right, while still touching $\mathcal{U}$. By linear programming, L intersects a vertex, which in this case is the spectrum locus vector for 625 nm and happens to fall right on the X-axis. The CMF thus peaks where we would expect it to, at the wavelength the response curve peaks at. A similar result holds for Y.

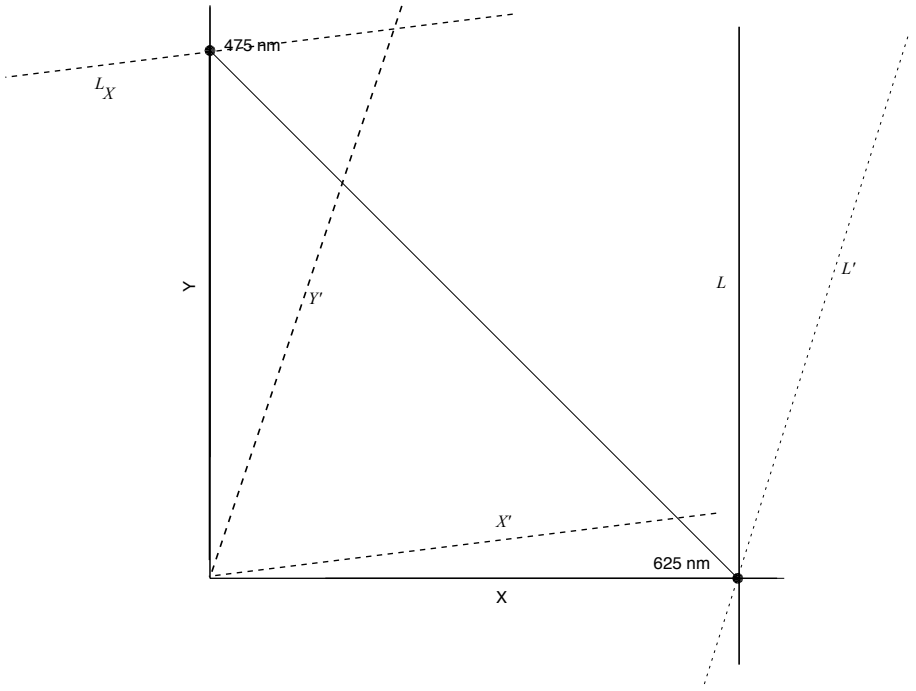

Figure 8: Changing One Primary

The discussion so far has been based on a natural set of primaries, but we will see geometrically that in this case the CMFs peak at the same wavelengths for any set of primaries. To begin with, suppose that we leave the primary X unchanged for now, but change the primary Y to some new primary Y'. In any bipartite field experiment, Y' would have to result from a real primary—i.e. one whose power at any wavelength is non-negative—so its spectrum locus vector would occur in the first quadrant, cutting through the triangle. (Note that such a Y' would actually be a polychromatic primary, because it stimulates both the left and the right cones.)

Visually, we would rotate the Y-axis in Figure 8 somewhat to the right, until it reaches some Y', as the figure shows. To maintain parallelism, the line L would rotate the same amount to the right, ending at some new line L', as shown. L' intersects the triangle as far to the right as possible, so its intersection point occurs at a vertex, and we can see from the figure that that vertex corresponds to 625 nm, just as it did previously. In fact, if Y were rotated any amount, while still cutting through the triangle, L' would still intersect the triangle at the spectrum locus vector corresponding to 625 nm, so the CMF for X would peak at 625 nm.

Now that Y has been rotated to some Y', we can choose a new second primary by rotating X to some X'. As before, X' must cut through the triangle $\mathcal{U}$ to insure a physically possible bipartite field experiment. We would draw the line $L_{X'}$ that is parallel to the X'-axis, and intersects the triangle as far as possible from that axis. We can see that, regardless of what X' is chosen, such an $L_{X'}$ would intersect the triangle at the vertex given by the spectrum locus vector for 475 nm, just where the first response curve peaks. No matter what new primaries are chosen, then, for X or for Y, the resulting CMFs will always peak at 475 and 625 nm.

Geometrically, this invariance of the CMF peaks depends on the *shape* of $\mathcal{U}$. The vertex at 625 nm (and the one at 475 nm) is very pointed, in the sense that a line through it could be rotated through a wide angle while still keeping $\mathcal{U}$ completely on its left. This fact allows us to adjust the Y-axis freely, and gives us complete latitude in choosing the corresponding primary; the X-axis can be similarly freely adjusted—without changing the CMF peaks. A subsequent example will show more rounded vertices, in which adjusting the primaries can alter the CMF peak wavelengths.

3.2.3 Two Overlapping Response Curves

The left side of Figure 9 shows another case of two response curves, but this time their peak wavelengths are 500 and 600 nm and they overlap slightly. As a result, the wavelengths in the center third of the spectrum stimulate both cones simultaneously, although the other wavelengths only stimulate one cone. The right side of the figure shows the set of unit-power colours, where the X-coordinate results from choosing a primary between 600 and 700 nm, and the the Y-coordinate results from choosing a primary between 400 and 500 nm. The triangle that occurred when there was no overlap has now expanded to something more like a quarter-circle. It can be seen, however, that the behavior with regard to CMF peaks is identical. The line L parallel to the Y-axis intersects $\mathcal{U}$ at the vertex corresponding to 600 nm, and, even if the Y-axis were rotated to any real primary Y' within $\mathcal{U}$, the new line L' would still intersect $\mathcal{U}$ at that vertex. The difference from the triangle is that L now intersects the vertex nearly tangentially, and that other intersections are at much narrower angles than previously.

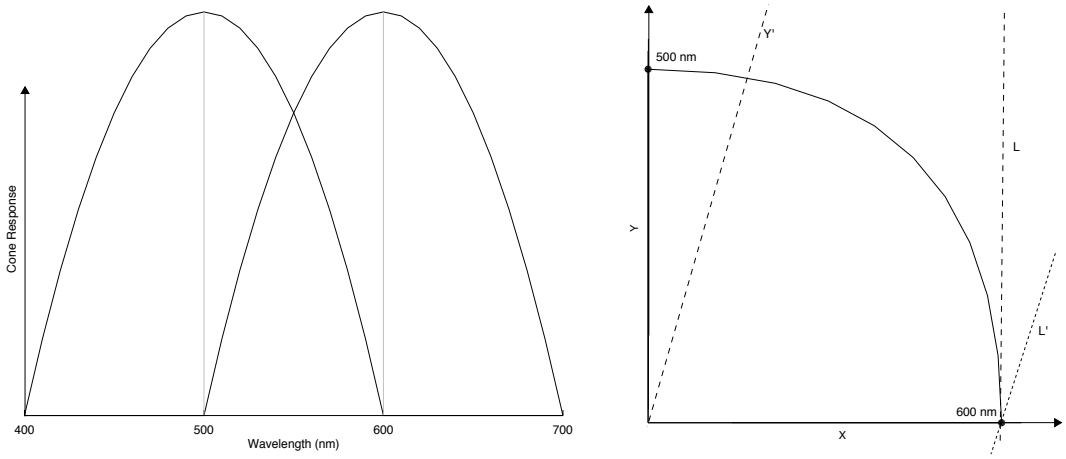

Figure 9: Response Curves and Unit-Power Colours for Slightly Overlapping Cones

Figure 10 shows a further example of two response curves, peaking at wavelengths 530 and 570 nm, but now with a much more significant overlap. Their unit-power hull is also significantly different. The right side of the figure shows a marked protuberance, and indicates the spectrum locus vectors for various wavelengths. The axes result from choosing primaries, one between 400 and 440 nm, and the other between 660 and 700 nm, that stimulate only one cone. In that basis, the line L given by the farthest parallel translation of Y intersects the unit-power hull at the vertex corresponding to 570 nm, so the CMF for X will peak

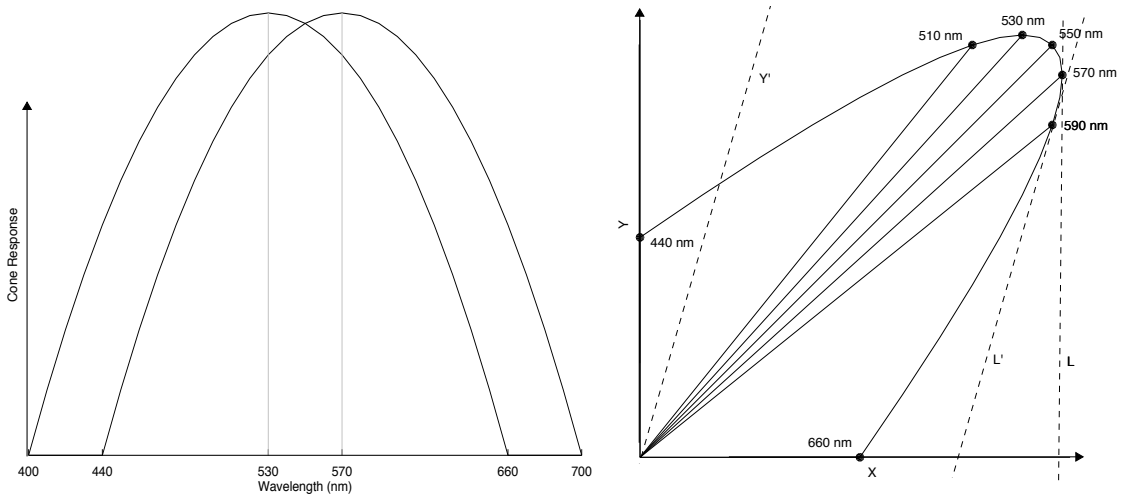

Figure 10: Response Curves and Unit-Power Colours for Significantly Overlapping Cones

there. Now suppose as before that choosing a different primary causes Y to be rotated to a new axis Y'. Then the corresponding L' goes not through 570 nm, but through 590 nm, as shown, and the CMF's peak wavelength shifts.

This example shows that CMF peaks, even when restricted to monochromatic primaries, need not be invariant. Furthermore, one could picture cases, where $\mathcal{U}$ is not too flat at a vertex but not too pointed there either, in which a peak wavelength varies as different sets of primaries are chosen, but only within a limited interval. The result would be a peak wavelength that was reasonably consistent in practice, but not perfectly so, which is just the behavior that the Thornton primaries show.

4 Application to the Thornton Primaries

4.1 CMFs for Different Bases

The previous section showed how the CMFs' peaks, and their variability as different sets of primaries are chosen, depend on the shape of the unit-power hull. The examples there illustrated the relevant geometric concepts in two dimensions. In this section, they will be applied to three-dimensional human colour space, and explain why the Thornton wavelengths recur consistently, but with some slight variation.

In three dimensions the convex set $\mathcal{P}$ of interest is the unit-power hull $\mathcal{U}$, which is a polyhedron rather than a polygon. In both two and three dimensions, a hyperplane is translated across a convex set; in two dimensions a hyperplane is a one-dimensional subspace generated by a single axis, while in three dimensions a hyperplane is a two-dimensional subspace, which is just an ordinary plane, generated by two of the axes.

The convexity of $\mathcal{U}$ follows by construction, because $\mathcal{U}$ is the convex hull of the spectrum locus vectors. Furthermore, standard convexity theory (Sect. 20 of Ref. 5) tells us that $\mathcal{U}$ is a polyhedron, and that each of its vertices occurs at the tip of a spectrum locus vector. The unit-power colours are a subset of the convex cone of all colours, which is itself a subset of

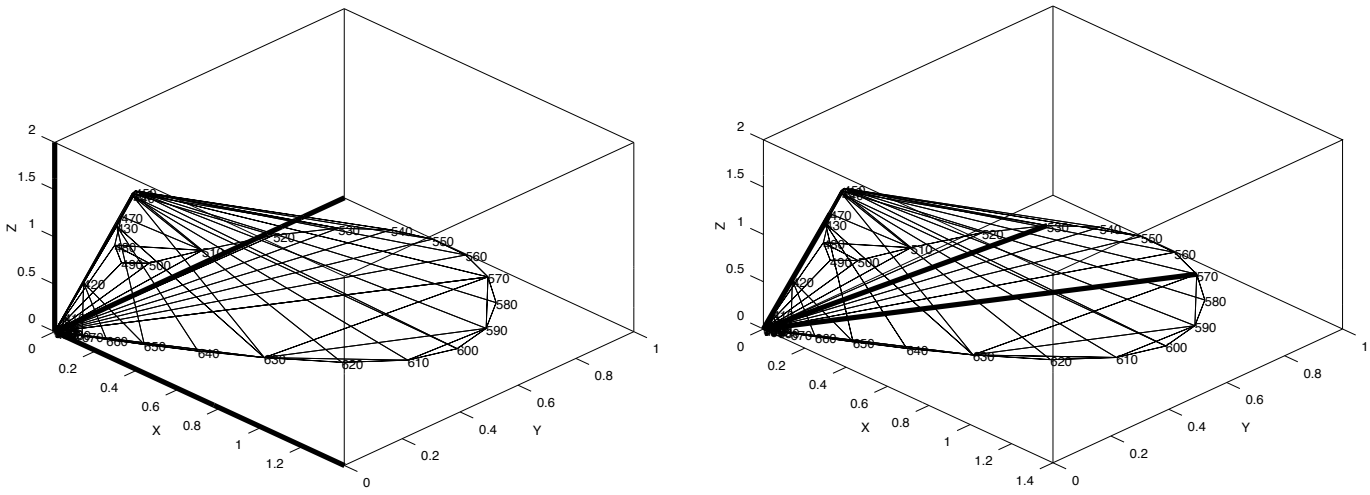

Figure 11: The Unit-Power Hull $\mathcal{U}$, with Two Different Bases

the three-dimensional vector space V.

Numerous bases can be chosen for V. A vector space and a fixed subset $\mathcal{S}$ can be visualized in two natural, mathematically equivalent ways. The first method is to draw the basis vectors as if they were orthogonal unit vectors in Euclidean space; each point in the subset $\mathcal{S}$ can be expressed as a linear combination of the basis vectors, so the entire subset can be drawn in this way. In this approach $\mathcal{S}$ takes on different shapes when the basis is changed. For instance, a sphere could be converted to an ellipsoid, or vice versa. The second method, which will be followed here, is to draw $\mathcal{S}$ so that its shape never changes, and then draw basis vectors where they would occur in relation to $\mathcal{S}$. A new trio of basis vectors would start at the origin, but point in different directions from the previous trio, while the subset itself would not look any different.

The left side of Figure 11 shows the unit-power hull $\mathcal{U}$, and highlights the three standard CIE basis vectors given by coordinates X, Y, and Z. They appear just as they would in a Euclidean space. Let us leave the subset $\mathcal{U}$ fixed, and suppose we are performing a colour-matching experiment, using monochromatic primaries at 470, 530, and 570 nm. The colour vectors of these primaries are then taken as a new basis for V. Since they result from real SPDs, all the new basis vectors must go through $\mathcal{U}$. Since they are monochromatic, they will lie along the spectrum locus vectors for 470, 530, and 570 nm, as shown in the right side of Figure 11. Note that $\mathcal{U}$ itself is identical in both parts of the figure—only the basis has changed.

Each basis produces its own set of colour-matching functions, which peak at some set of wavelengths. Let us start with the CIE 1931 basis, and apply the results of the previous section. To find where the X-coordinate peaks, we parallel-translate the Y-Z-plane as far as possible to the right, while still keeping in contact with $\mathcal{U}$. The projected view in the bottom right corner of Figure 2 makes the operation clear; since the Z-axis is pointing directly out of the page, we need only translate the Y-axis as far as possible to the right, ending up at the dotted line. This line intersects $\mathcal{U}$ at the vertex corresponding to 600 nm, and we see in Figure 1 that the CMF $\bar{x}$ corresponding to X does indeed peak at 600 nm. The peaks for Y

and Z can be determined similarly to be 450 and 555 (midway between 550 and 560) nm, agreeing with Figure 1.

Now suppose that we are finding peaks of CMFs for the basis shown in the right half of Figure 11. To find the peak for the coordinate corresponding to the 570 nm primary, we construct the plane generated by the 470 and 530 nm primaries. It can be seen from the figure that this plane is not vertical in the sense that the Y-Z-plane is vertical, but is rather tilted and rotated somewhat. Nevertheless, it can be translated across $\mathcal{U}$ until it just touches a single vertex, which in this case occurs at the tip of the spectrum locus vector for 600 nm. Figure 12 shows the actual colour-matching functions, and the peak for the 570 nm CMF does indeed occur at 600 nm. The other two peaks can be derived similarly, and occur at 450 and 530 nm.

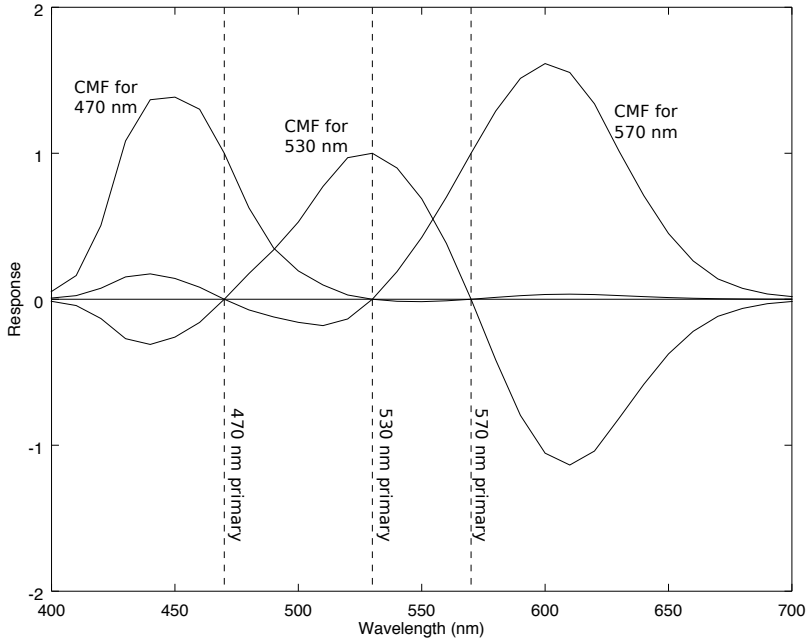

Figure 12: Colour-Matching Functions for the Primaries 470, 530, and 570 nm

The CMF peaks for the CIE basis and the new basis are nearly identical; the only difference occurs for the middle peak, which is at 555 nm in one case and 530 nm in the other—and considering the relatively crude 10 nm bandwidth, a difference of 25 nm might not be very much. As early as 1953, MacAdam[6] had observed that the CMF peaks seemed "amazingly similar," regardless of the primaries chosen. Thornton[3] estimated the peaks by choosing a representative sample of 792 bases (each with the first primary between 410 and 580 nm, the second between 490 and 570 nm, and the third between 580 and 680 nm), and averaging the peaks of the resulting CMFs. The results (for the 1931 Standard Observer) were the Thornton primaries of 447, 541, and 604 nm. Furthermore, the average deviations for the peak wavelengths were very small, at 0.4, 2.4, and 3.9 nm respectively (see Table 1 of Ref. 3). His results therefore demonstrate quantitatively that CMF peaks show considerable consistency. The rest of the paper will show the geometry origins of this consistency.

4.2 Geometric Explanation

The geometric expositions of our previous examples explain both the consistency of the Thornton primaries, and the fact that that consistency is not perfect. The main conclusion is that the shape of $\mathcal{U}$, and in particular the pointedness of its vertices, is responsible for both the approximate consistency and for the degree of variability.

The most noticeable feature of $\mathcal{U}$, as seen in Figure 2, is the sharp vertex, like the apex of a pyramid, appearing at 450 nm. This feature recalls the vertices seen in the triangle in Figure 7. In both cases a large variety of hyperplanes, at a wide range of orientations, reach their maximum point in $\mathcal{U}$ at a single vertex. The hyperplane of interest in the basis in the right side of Figure 11, for example, is parallel to the plane spanned by the 530 and 570 nm primaries, both of which are nearly in the CIE X-Y-plane. Even if these two primaries were changed significantly, as long as they both remain above about 480 nm and separated by at least 20 nm or so (two very natural conditions in practice), the plane they generate will not be tilted enough from the plane through 530 and 570 nm to move the farthest intersection point with $\mathcal{U}$ from the 450 nm vertex. As a result, Thornton found an average change of only 0.4 nm over 792 cases, so that the 450 nm peak is effectively constant for practical purposes. Similar arguments apply, though not as strongly, to the peaks at 540 and 600 nm. Rather than a sharp apex, the spectrum locus vectors for the latter two values show sharp edges, which curve away from the vertices gradually, rather than exhibiting the sharp point seen at 450 nm. The gradual curving away is similar to the behavior seen in the right side of Figure 10, and unlike the behavior in the right side of Figure 7. A small adjustment in the touching hyperplane, then, changes one of these vertices more than the 450 nm vertex (2.4 or 3.9 nm vs. 0.4), but the two vertices are still sharp enough that any likely changes are limited.

Figure 3 unifies the arguments for all three peaks. That figure shows the Thornton tetrahedron $\mathcal{T}$, formed by the spectrum locus vectors of the Thornton primaries T_1, T_2, and T_3. Three faces of the tetrahedron touch the origin. Typically, primaries are chosen to be fairly evenly spaced within the visible spectrum. Furthermore, since the Z-component is very near 0 in the right half of the spectrum (see Figure 1), the spectrum locus vectors from 550 to 700 nm lie almost perfectly within the X-Y-plane, and $\mathcal{U}$ is almost flat there.

As a result, the plane spanned by the two primaries of longest wavelengths is not too different from the 0-T_2-T_3 face of the tetrahedron. Similarly, since the primary of shortest wavelength is usually less than 500 nm, the plane spanned by the spectrum locus vectors of the two shorter wavelength primaries is not tilted too much from the 0-T_1-T_2 tetrahedral face, and the remaining plane, spanned by the primaries of shortest and longest wavelengths, is not tilted too much from the 0-T_1-T_3 face. When any face of $\mathcal{T}$ is translated across $\mathcal{U}$, is should reach its maximum at the opposing vertex of the tetrahedron—and all three of those vertices occur as Thornton primaries. Since the vertex at 450 nm is extremely pointed, and the other two are moderately sharp, the maximizing vertices do not change much. In this way, the CMF peak wavelengths are constrained to three narrow regions of the spectrum, over a wide range of sets of primaries, which is just the behavior observed empirically.

To sum up the geometric arguments, the consistency of the CMF peak wavelengths depends on two facts:

1. Typical primaries are monochromatic and widely spaced across the visible spectrum, and
2. The Thornton tetrahedron $\mathcal{T}$ is a good approximation to the unit-power hull $\mathcal{U}$.

The first fact results from practical considerations. A set of closely spaced primaries such as 490, 500, and 510 nm, for instance, produces peak wavelengths at 440, 500, and 600 nm, missing the second Thornton primary by 40 nm—such a choice would be very unusual in practice, however, because most experimenters would tend to cover the spectrum more evenly. Using monochromatic primaries insures that the corresponding basis vectors lie on the boundary surface of $\mathcal{U}$, while a basis vector for a polychromatic primary would go through the interior of $\mathcal{U}$. If the basis vectors travelled through the interior, the resulting planes could easily deviate significantly from the Thornton tetrahedron, while vectors on the boundary will produce planes that are much closer to the faces of $\mathcal{T}$.

Another practical restriction is that the chosen primaries must be real, so that they could actually be produced for a colour-matching experiment. The middle peak of the CIE CMFs, shown in Figure 1, occurs between 550 and 560 nm, about 15 nm to the right of the Thornton primary of 540 nm, six times larger than Thornton's average deviation of 2.4 nm. The CIE primaries, though, are imaginary, so their spectrum locus vectors only intersect $\mathcal{U}$ at the origin; they are otherwise completely exterior. Any real primaries, however, must travel through some subset of $\mathcal{U}$ for some interval before exiting. The argument that planes generated by pairs of primaries are close to the faces of the Thornton tetrahedron require those primaries to be contained in $\mathcal{U}$, so they must be physically possible SPDs rather than mathematical fictions.

The second fact results from the geometry of colour perception. An exhaustive calculation shows that $\mathcal{T}$ has a larger volume (as a percentage of $\mathcal{U}$) than any other tetrahedron formed similarly from three spectrum locus vectors. Heuristically, a volume-maximizing tetrahedron in a convex set should approximate that set's shape as well as possible, or else moving one or more vertices of the tetrahedron around in the set could increase the tetrahedron's volume. The vertices of such a volume-maximizing approximation would be expected to fall along any extreme protrusions of the convex set being approximated. In this case, of course, the tetrahedron correctly identifies the pointed protuberance at 450 nm. Overall, then $\mathcal{T}$ is a good proxy for empirical colour-matching data, so it is not surprising that its vertices show up consistently as the Thornton primaries.

4.3 Relation to Some Previous Work

Our constructions can also be used to demonstrate geometrically some results that had previously been proved algebraically. One result is due to Brill and Worthey[7] in 2006, and another to MacAdam[6] in 1953. Both these results assert—very counterintuitively—that the CMF for a primary depends very little on that primary itself, but depends very much on the other two primaries. In fact, once two primaries are fixed, the shape of the CMF for the third primary is determined, and specifying the wavelength for that primary's wavelength only scales that shape by some constant multiple. The geometric reasoning of the current paper makes these assertions intuitively clear.

In some online notes, Brill and Worthey use matrix manipulations and Cramer's rule to prove the following result: Suppose there are three monochromatic primaries (call them λ_1, λ_2, and λ_3) with corresponding CMFs $\bar{x}_1(\lambda)$, $\bar{x}_2(\lambda)$, and $\bar{x}_3(\lambda)$; if λ_1 is changed to a different wavelength λ_1', but λ_2 and λ_3 remain the same, then the CMF for λ_1' is a scalar multiple of $\bar{x}_1(\lambda)$. This same result can be seen geometrically. The values of the CMF for the first

primary are determined by translating the plane generated by the other two primaries. To find the value of the first CMF at some wavelength λ, translate that plane until it touches the tip p of the spectrum locus vector for λ, as in Figure 4, and project onto the appropriate axis. When all the translations for all the wavelengths are made, the result will be a three-dimensional version of Figure 5, with $\mathcal{P}$ replaced by $\mathcal{U}$ and the lines replaced by planes. While the magnitudes of the coordinates along the x-axis will change if x is replaced by some new vector x', the ratios of the coordinates remain identical. Since the coordinates along a primary's axis are just the CMF values for that primary, the CMFs resulting from x and x' will be scalar multiples, as Brill and Worthey asserted.

MacAdam's result is similar. As the first line of his abstract says: "The shape of the color-mixture curve for any primary is completely determined by the choice of the line in the chromaticity diagram representing the additive mixture of the other two primaries." Once one recalls that the chromaticity diagram lies in the affine subspace $X+Y+Z=1$ of V, one sees that the linear span of the origin and a line in the chromaticity diagram is a plane P in V. Since this line represents "the additive mixture of the other two primaries," the plane P itself is the linear span of those two primaries, so it is just the plane that we have been parallel-translating to evaluate the peak wavelength of the CMF for the first primary, and, considering Figure 5, the value of that CMF at any wavelength. The parallel translation, and the relative spacing of the projections onto the first primary, do not actually depend on the direction of the first primary, which can only assign relative values that change the size of its CMF but keep its shape the same.

5 Summary

This paper has explained geometrically why CMFs tend to peak around the Thornton primaries in practical applications. The unit-power hull $\mathcal{U}$, a convex polyhedron which encodes the results of colour-matching experiments, was constructed as the convex hull of spectrum locus vectors. Choosing three primaries for such an experiment is equivalent to choosing a basis of the three-dimensional colour space V, where $\mathcal{U}$ lives. Each primary has an associated CMF, whose peak wavelength is the solution to a linear programming problem: the wavelength corresponding to the vertex on $\mathcal{U}$ at which that primary attains a maximum value is that primary's peak wavelength. Surprisingly, the solution to this problem depends more on the other two primaries than it does on the primary of interest. The basis vectors for the other two primaries generate a plane in V, which can be parallel-translated over $\mathcal{U}$ to the farthest vertex possible, whose wavelength is the first primary's peak.

These constructions show that peak wavelengths depend on the *shape* of $\mathcal{U}$. That shape is approximated well by the Thornton tetrahedron $\mathcal{T}$, which is the pyramid formed by the colour vectors for the Thornton primaries. As long as an experimenter chooses three monochromatic primaries that are fairly widely spaced across the visible spectrum, the planes mentioned in the previous paragraph will be approximately parallel to the faces of $\mathcal{T}$, and this fact, combined with the sufficient pointedness of $\mathcal{T}$'s vertices, insures that those vertices will not change much as the experimenter's primaries change, explaining the consistent CMF peaks that have been observed in practice.

References

1. Paul Centore, *The Geometry of Colour*, 221 pages, 2017.
2. Deane B. Judd, "The 1931 I. C. I. Standard Observer and Coordinate System for Colorimetry," JOSA, Vol. 23, pp. 359-374, October 1933.
3. William A. Thornton, "Spectral Sensitivities of the Normal Human Visual System, Color-Matching Functions and Their Principles, and How and Why the Two Sets Should Coincide," Color Research & Application, Vol. 24, Number 2, April 1999, pp. 139-156.
4. Bernard Kolman & Robert E. Beck, *Elementary Linear Programming with Applications, 2nd ed.*, Academic Press, 1995.
5. S. R. Lay, *Convex Sets and Their Applications*, Dover Publications, 2007.
6. David L. MacAdam. "Dependence of Color-Mixture Functions on Choice of Primaries," JOSA, Vol. 43, Issue 6, 1953, pp. 533-538.
7. Michael H. Brill & James A. Worthey, "Color Matching Functions When One Primary Wavelength is Changed," Color Research & Application, Vol. 32, No. 1, pp. 22-24, February 2007.

