The Unit-Power Hull for Minimal-Power Metamers

Paul Centore

Abstract

The same colour perception can be produced by many different spectral power distributions (SPDs), called metamers. This paper shows geometrically that the minimal-power metamer for a target colour has power at only two wavelengths. In three-dimensional colour space, the spectrum locus vectors are the colours that arise from unit-power, single-wavelength SPDs. Their convex hull, which we will call the unit-power hull U, consists of all colours that can be produced with unit power or less. An equivalent problem to finding a minimal-power metamer is finding the SPD at which a given chromaticity achieves a maximum luminance, using only one unit of power. The colours of one chromaticity lie on a ray from the origin and achieve maximum luminance (for one unit of power) at the point p where the ray exits U through a triangular face. With a discretized spectrum, convexity arguments imply that the minimal-power SPD for p has power at only three wavelengths, one for each of the triangle's vertices. As the discretization interval goes to 0, two of the wavelengths asymptotically approach each other, meeting ultimately at a single wavelength, so that a continuous minimal-power metamer has power at only two wavelengths.

Index Terms: metamer, unit-power hull, spectrum locus, convex hull, minimal power

1 Introduction

A color perception occurs when a physical spectral power distribution (SPD) enters the eye. An SPD can be expressed as a function $s(\lambda)$, where λ is a wavelength in the physical spectrum, between 400 and 700 nm. $s(\lambda)$ gives the physical power, typically expressed in Watts, of a colour stimulus at the wavelength λ. In a classical colour-matching experiment, an observer's field of view is restricted to two small, adjacent regions, one of which emits one SPD, and the other of which emits another SPD. Even though the two SPDs might be very different physically, a human can perceive their colours as identical. Two such SPDs are called *metamers*.

Though they produce identical colours, two metameric SPDs could themselves require very different total power levels over the visible spectrum. A practical problem, especially for mobile devices with limited battery life, is producing a desired colour with as little power as possible, so it would be helpful to identify a minimal-power metamer for a given target colour. This paper use some geometric constructions to show:

1. If SPDs are treated as continuous functions, then a minimal-power metamer has non-zero power at no more than two wavelengths (and zero power elsewhere), and
2. If SPDs are treated as discrete functions, then a minimal-power metamer has non-zero power at no more than three wavelengths (and zero power elsewhere).

(Apart from special cases, such as a colour produced by a single-wavelength SPD, continuous and discrete minima have power at *exactly* two or three wavelengths, respectively.)

SPDs can be expressed as either continuous or, more practically, discrete functions. A discretization divides the visible spectrum into channels of some width; a typical choice is 31 channels, each 10 nm wide, centered on the wavelengths 400, 410, 420,...,690, 700 nm. As the discretization of the spectrum becomes increasingly fine, the channel widths go to 0 nm, and a discrete SPD becomes a continuous SPD. Simultaneously, two of the three wavelengths in the discrete minimal-power SPD approach each other, converging in the limit to a single wavelength, so that the continuous minimal-power SPD only has two wavelengths.

By allowing the mathematical fiction of SPDs with one or more negative power levels, called *imaginary SPDs*, the set $\mathcal{S}$ of SPDs becomes a vector space. The power of an SPD is found by integrating the power in each wavelength over the visible spectrum; for a discrete SPD, integration reduces to simply summing the powers in the wavelength channels. In either case, power is a linear functional P on $\mathcal{S}$. Metamerism defines an equivalence relationship on SPDs: two SPDs are equivalent if and only if they produce identical colour perceptions. In fact, we can *define* a colour as an equivalence class of SPDs, and we can further define a transformation T on $\mathcal{S}$ that sends an SPD $s(\lambda)$ to the colour perception it produces. Grassmann's laws[1] imply that T is a linear transformation, so a colour is an element in a vector space $\mathcal{V}$, sometimes called *colour space*. Not every vector in $\mathcal{V}$ is a real colour; some vectors, called *imaginary colours*, do not correspond to any visual perception.

Empirical investigations have shown that V has three dimensions. In 1931, the Commission Internationale de l'Éclairage (CIE) assigned[2] $\mathcal{V}$ a basis, whose coordinates are denoted X, Y, and Z. A colour v in $\mathcal{V}$ can result from a wide variety of metameric SPDs: any $s(\lambda) \in \mathcal{S}$ such that $T(s) = v$ will do. By Grassmann's laws, adding two colours v_1 and v_2 in $\mathcal{V}$ is equivalent to choosing two SPDs $s_1(\lambda)$ and $s_2(\lambda)$, such that $T(s_1) = v_1$ and $T(s_2) = v_2$, and adding the SPDs instead:

$$v_1 + v_2 = T(s_1 + s_2). \tag{1}$$

This definition is independent of which SPDs are chosen as representatives. Similarly, a colour can be multiplied by a scalar k to produce a new colour in $\mathcal{V}$, or equivalently, any representative SPD for v could be multiplied by k.

Each ray emanating from the origin in $\mathcal{V}$ consists of colours that are all scalar multiples of one another. The orientation of a ray in $\mathcal{V}$ is expressed by two *chromaticity coordinates* x and y :

$$x = \frac{X}{X+Y+Z}, \tag{2}$$

$$y = \frac{Y}{X+Y+Z}. \tag{3}$$

The *chromaticity diagram* $\mathcal{C}$ is the projection of $\mathcal{V}$ (or at least of all the non-imaginary colours in $\mathcal{V}$) onto the plane $X + Y + Z = 1$, where each $v \in \mathcal{V}$ is mapped to the point where the chromaticity ray through v intersects $X + Y + Z = 1$. x and y can serve as coordinates for that plane. Equations (2) and (3) define a chromaticity for each colour in $\mathcal{V}$.

The coordinate Y is called *luminance*, and was deliberately chosen by the CIE to represent how light or dark a colour is. Together, the chromaticity coordinates and luminance define another coordinate system for $\mathcal{V}$, one which fits well with human perception. As one moves along a ray, away from the origin, the colours increase in luminance (they become brighter), but their chromaticity (which involves hue and saturation) remains the same. Turning up a rheostat on a red warning light, for instance, would increase the output SPD by a scalar multiple; the light would appear to be the same red, but brighter.

In a classical colour-matching experiment, one half of a small circle, which is the only visual stimulus for an observer, displays the unit-power, monochromatic (i.e. all its power is concentrated in a single wavelength) SPD $\mathbf{1}_\lambda$ for some wavelength λ. Three other monochromatic SPDs, called primaries, are available for an observer to superpose, varying their power levels at will. Typically, two primaries are combined in the second half of the circle, and the third primary is transferred to the first half of the circle. The observer adjusts the primaries' power levels until the two halves of the circle have identical colours. The chosen power levels are recorded, with the transferred primary being assigned a negative power level. The observer steps through individual target wavelengths, usually 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 λ, the power level required of that primary in the colour match involving $\mathbf{1}_\lambda$. A colour-matching experiment thus results in three CMFs, all of which are discrete functions over the visible spectrum.

The three CMF values at a wavelength λ can be re-expressed as the components of a three-dimensional *spectrum locus vector*. Figure 1 plots those figures in $\mathcal{V}$. Joining the tips of the vectors together sequentially forms a curve that we will call the *3d spectrum locus*. (An earlier source[3] used the term *locus of unit monochromats* (LUM).) Each spectrum locus vector corresponds to a unit-power monochromatic SPD $\mathbf{1}_\lambda$, and any SPD of unit power can be written in $\mathcal{S}$ as a convex linear combination of unit-power monochromatic SPDs. The corresponding construction in $\mathcal{V}$ is the convex hull of the 3d spectrum locus (augmented by the origin), which will be called the *unit-power hull* $\mathcal{U}$. It is shown in Figure 2. It consists of all colours that can be produced with an SPD of unit power or less. Note that although any colour in $\mathcal{U}$ *can* be produced with no more than unit power, those colours also have many metamers that use considerably more power.

Mathematical analysis shows that the choice of primaries in a colour-matching experiment is incidental: the CMFs for one set of primaries can always be transformed into the CMFs for a second set of primaries. Choosing a new set of primaries is equivalent to choosing a new basis for $\mathcal{V}$. Figure 2 shows $\mathcal{U}$ with the CIE coordinates X, Y, and Z used as basis vectors. Even if another basis were used in place of X, Y, and Z, though, the relationships that define $\mathcal{U}$ would not change. The unit-power hull in any coordinate system will always be convex, and have the same vertex-edge-face structure. For instance, the tips of the spectrum locus vectors for 450, 560, and 570 nm form a triangular face of $\mathcal{U}$. Those three wavelengths will always form a triangular face of the unit-power hull regardless of the basis chosen for $\mathcal{V}$.

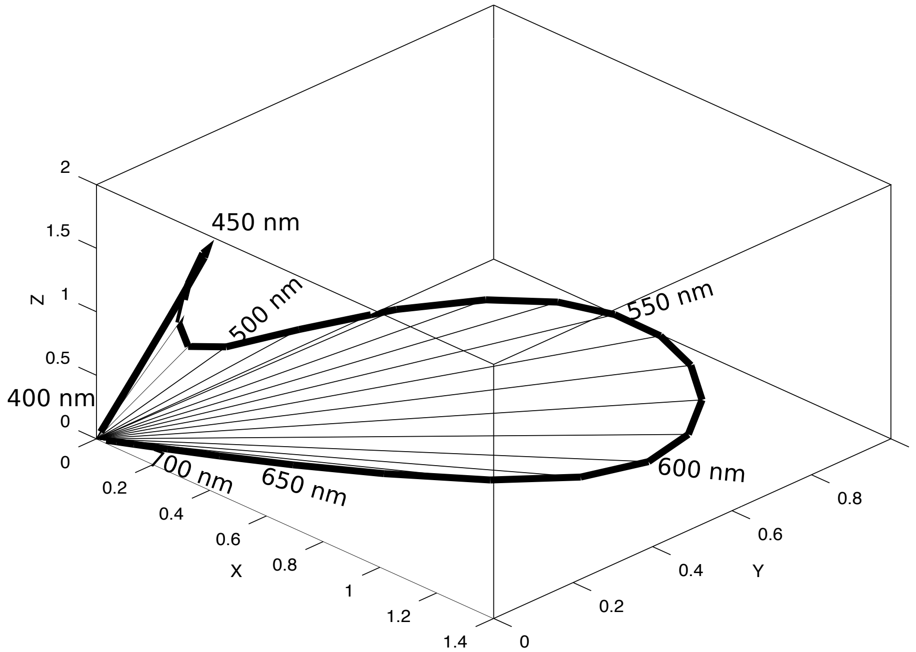

Figure 1: The Spectrum Locus Vectors and the 3d Spectrum Locus

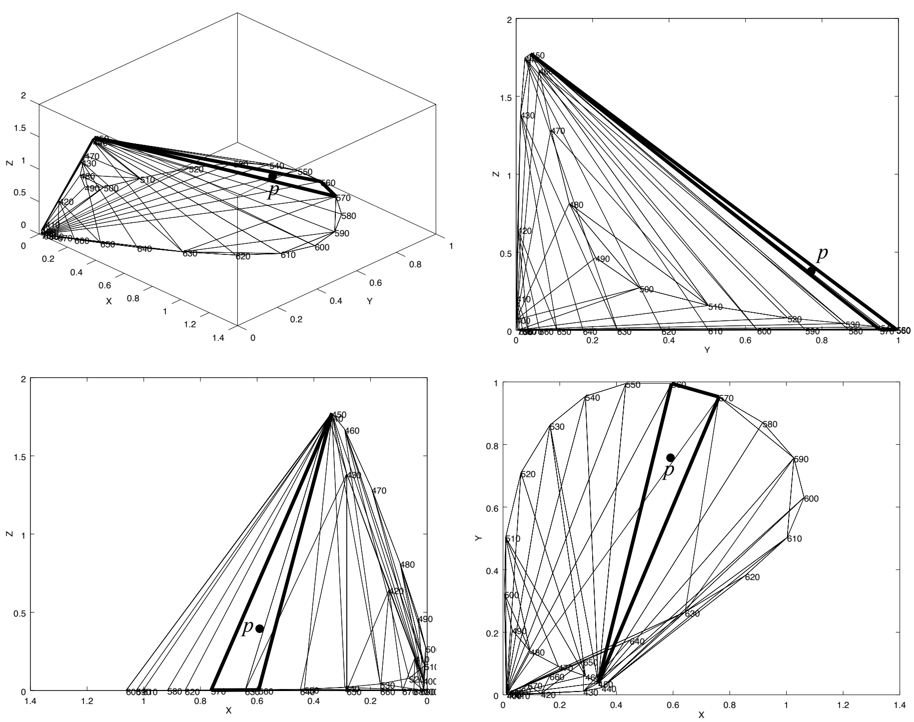

Figure 2: The Unit-Power Hull $\mathcal{U}$ (Top Right: View along X-axis; Bottom Left: View along Y-axis; Bottom Right: View along Z-axis)

The foregoing constructions have laid the groundwork for investigating the minimal-power metamer problem:

Given a target colour v, find a minimal-power metamer, that is, an SPD s_{min} that produces v with as little power as possible.

Rather than attack this problem directly, our geometric constructions will solve a problem that is mathematically equivalent:

Find an SPD s_{max} that uses no more than one unit of power, that produces a colour with the same chromaticity as v, but with maximum luminance.

We will see that s_{min} is a scalar multiple of s_{max}.

To solve the second problem geometrically, consider that all the colours with the same chromaticity as v fall on the ray R_v, that starts at the origin and goes through v. The farther a colour on R_v is from the origin, the greater its luminance. Since $\mathcal{U}$ contains all colours that can be produced with unit power or less, we would like to move along R_v as far from the origin as possible, while still staying in $\mathcal{U}$. The colour v_{max} of maximum luminance will therefore occur at the point p where R_v exits $\mathcal{U}$.

While v_{max} has many metamers, with many different power levels, the geometric construction implies that there is a unique metamer of unit power. Every bounding face of $\mathcal{U}$ is triangular, and its vertices occur at the tips of the spectrum locus vectors of three wavelengths; call them λ_1, λ_2, and λ_3. Since p is on the boundary, convexity theory implies that it can occur only as a convex combination of those three vertices. Furthermore, the coefficients in that combination are just the barycentric coordinates β_1, β_2, and β_3, of p in that face, and those coefficients are unique (Thm. 2.25 & 2.26 of Ref. 4). It follows that the unique SPD of unit power for p really is s_{max}, and is given by

$$s_{\text{max}}(\lambda) = \begin{cases} \beta_1, & \text{if } \lambda = \lambda_1, \\ \beta_2, & \text{if } \lambda = \lambda_2, \\ \beta_3, & \text{if } \lambda = \lambda_3, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Since barycentric coordinates sum to 1, the power levels in Expression (4) also sum to 1, so s_{max} maximizes luminance with unit power, as desired. To produce s_{min}, just multiply s_{max} by the ratio of the luminance of v to the luminance of v_{max}.

Since the input to the second problem is just a chromaticity, rather than a complete colour, all the solutions can be conveniently displayed in the chromaticity diagram $\mathcal{C}$. Every triangular face of $\mathcal{U}$ is projected to a triangle in $\mathcal{C}$, and this projection preserves barycentric coordinates. Solutions only occur on the triangles not touching the origin, and Figure 3 shows the projections of all these triangles. The chromaticity ray R_v, including p, is projected to its chromaticity in $\mathcal{C}$, so there is no confusion when Figure 3 reuses p to denote that chromaticity. p is in some triangle τ, whose vertices occur at three wavelengths (450, 560, and 570 nm in this case). The minimal-power SPD for p uses the barycentric coordinates of p in τ, exactly as in Expression (4). Therefore, apart from special cases like single-wavelength colours, minimal SPDs in the discrete case have exactly three non-zero wavelengths.

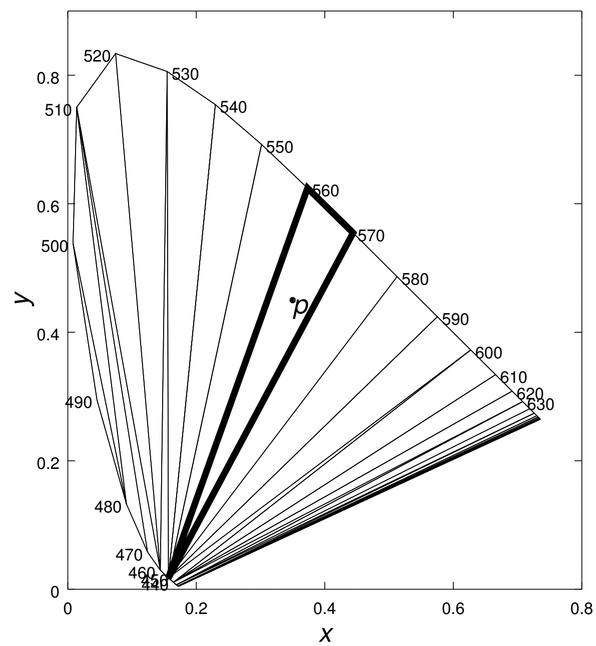

Figure 3: Projections of Bounding Triangles of $\mathcal{U}$ onto the Chromaticity Diagram

The transition from discrete SPDs to continuous ones is natural. Decrease the wavelength interval of 10 nm in our discrete example to 5 nm, then 1 nm, then 0.1 nm, and so on, getting arbitrarily close to 0, which is the continuous case. As the wavelength interval decreases by a factor of k, Figure 3 will have k times as many triangles, but each triangle will be k times as thin. Each triangle has a clear apex, like 450 nm in our example, that will not change much as the interval shrinks. Each triangle also has a clear base, consisting of two vertices of adjacent wavelengths, like 560 and 570 nm in our example. As the wavelength interval decreases, these two vertices will move steadily closer to each other. In the limit, the triangle collapses to a line segment, which contains p. One end of the segment is at a wavelength λ_{1c}, which should be near the original apex wavelength, and the other end is at a wavelength λ_{2c}, somewhere between the two wavelengths bounding the base. p is thus a convex combination of λ_{1c} and λ_{2c}, so, analogously to Expression 4, the unique continuous minimal-power SPD for p will have non-zero power at only two wavelengths.

The paper is organized as follows. First, some well-established colour science constructions are summarized. These constructions are used in the second step, which formulates the minimal-power metamer in two equivalent ways. The second formulation lends itself to a novel geometric interpretation, which is this paper's main contribution. A new object, the unit-power hull, is constructed. Third, the geometry is developed in the discrete case, culminating in the conclusion that a discrete minimal-power metamer has power at only three wavelengths. Fourth, the two formulations are shown to provide complementary advantages: the first is useful for computations, and the second for understanding. Fifth, the fact that continuous minimal-power SPDs need power at only two wavelengths is derived as a limiting case of discrete SPDs, as the wavelength interval shrinks to 0. Finally, the methods of this paper are compared with previous work on the minimal-power metamer problem, and a summary is given.

2 Colour Science Constructions

2.1 The Vector Space of SPDs

The human visual 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. While continuous in reality, in applications $s(\lambda)$ is often discretized into a set of 31 values at increments of 10 nm from 400 to 700 nm. A finer discretization, with more wavelengths in its domain, can be achieved by using a smaller increment. In the limit, as the increment goes to 0, the discrete case approaches the continuous case. In this paper, we will first reach conclusions in the discrete case, with the standard increment of 10 nm, and then let the increment go to 0. As it does, various constructions will be modified, and the form of the conclusions in the continuous case will become clear.

In either the continuous or the discrete case, the set $\mathcal{S}$ of SPDs can be assigned a natural addition (two SPDs $s_1(\lambda)$ and $s_2(\lambda)$ can be added by defining $(s_1 + s_2)(\lambda) = s_1(\lambda) + s_2(\lambda)$ for each wavelength λ) and scalar multiplication (k times an SPD $s(\lambda)$ is just $ks(\lambda)$). While power levels must be non-negative at each wavelength for a physically possible SPD, it is useful to allow the mathematical fiction of negative power levels at some or all wavelengths. An SPD with one or more negative power levels will be called *imaginary*. If we include imaginary SPDs, then the set $\mathcal{S}$ becomes a vector space.

$\mathcal{S}$ is infinite-dimensional in the continuous case and finite-dimensional in the discrete case. An SPD is called *monochromatic* if all its power is restricted to a single wavelength. The monochromatic SPD of unit power at wavelength λ will be denoted $\mathbf{1}_\lambda$. In the discrete case, the unit-power monochromatic SPDs define a convenient basis for $\mathcal{S}$. The standard increment of 10 nm thus results in 31 basis vectors, so $\mathcal{S}$ is 31-dimensional in this discretization.

The power, or total power, of an SPD $s(\lambda)$ is given in the continuous case by integrating $s(\lambda)$ over the set of all wavelengths from 400 to 700 nm, and in the discrete case by summing up the power levels $s(\lambda)$ over the wavelengths λ included in the discretization; in the standard 10 nm increment case, there will be 31 power levels, whose sum gives the total power of s. Whether continuous or discrete, the power of an SPD can be seen as a linear functional P on the vector space $\mathcal{S}$.

2.2 The Vector Space of Colours

In a classical colour-matching experiment, two different SPDs are displayed side-by-side with no other stimulus in view, and an observer adjusts them until their colours appear identical. Notably, very different SPDs can produce exactly the same colour. Two different SPDs that appear identical are called *metamers*, or said to be *metameric*. A *colour* in fact, can be defined as an equivalence class of SPDs that produce an identical perception, so a colour can be seen as a maximal set of metamers. Every real SPD produces some colour perception, so any SPD can be assigned unambiguously to a colour. Furthermore, Grassmann's laws (Sect. 4.3.2 of Ref. 1) imply that this assignment is a linear transformation T, so the set of all colours must be a subset of a vector space $\mathcal{V}$, and empirical research has determined that $\mathcal{V}$, which is sometimes called *colour space*, is three-dimensional. As a linear transformation,

T unambiguously maps any SPD s in $\mathcal{S}$ to a three-dimensional colour vector v in $\mathcal{V}$, i.e. $T(s) = v$.

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

$$X(s) = \sum_{\lambda=400 \text{ nm}}^{700 \text{ nm}} s(\lambda)\bar{x}(\lambda), \tag{5}$$

$$Y(s) = \sum_{\lambda=400 \text{ nm}}^{700 \text{ nm}} s(\lambda)\bar{y}(\lambda), \tag{6}$$

$$Z(s) = \sum_{\lambda=400 \text{ nm}}^{700 \text{ nm}} s(\lambda)\bar{z}(\lambda), \tag{7}$$

where $\bar{x}$, $\bar{y}$, and $\bar{z}$ are colour-matching functions (CMFs) that the CIE also standardized. In the CIE coordinate system, the linear transformation $T : \mathcal{S} \to \mathcal{V}$ is written

$$T(s) = [X(s), Y(s), Z(s)]\,. \tag{8}$$

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 real (discrete) SPD is a positive linear combination of monochromatic SPDs, and since $\mathcal{S}$ is mapped into $\mathcal{V}$ by a linear transformation, the colours in $\mathcal{V}$ that can be physically produced are non-negative linear combinations of the colours resulting from the basis of unit-power monochromatic SPDs, each of which has CIE coordinates

$$T(\mathbf{1}_\lambda) = [X(\mathbf{1}_\lambda), Y(\mathbf{1}_\lambda), Z(\mathbf{1}_\lambda)] = [\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)] \tag{9}$$

for some wavelength λ. These vectors, the images under T of the unit-power monochromatic SPDs, are called *spectrum locus vectors*, and the curve their tips trace out will be called the *3d spectrum locus*. See Figure 1 for an illustration. (Some earlier work[3] used the term *locus of unit monochromats (LUM)* for what we are calling the 3d spectrum locus.) As wavelengths move to either end 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.

Since the linear transformation $T : \mathcal{S} \to \mathcal{V}$ is determined by its action on a basis of $\mathcal{S}$, and since the unit-power monochromatic SPDs form a basis for $\mathcal{S}$, the 3d spectrum locus in $\mathcal{V}$, which is their image under T, in some sense contains all the information about colour, and thus figures in many colour constructions. For example, the set of all non-negative linear combinations, also called the *convex cone*, of the spectrum locus vectors is just the set of all physically possible colours. A later section will develop a further interesting structure, called the unit-power hull.

The CIE basis for $\mathcal{V}$ is convenient and customary, but also arbitrary. Any other set of three linearly independent vectors could serve equally well as a basis. Any structure of interest, almost by definition, does not depend on the choice of basis. The convex cone, for instance, would be defined identically in any other basis, and would still be identical with the set of all physically possible colours. It expresses an invariant relationship in that any colour can be written in terms of single-wavelength SPDs.

2.3 Chromaticity and Luminance

While the linear algebraic approach assigns colours coordinates in terms of a basis, chromaticity and luminance are more perceptually relevant coordinates.

Loosely speaking, luminance refers to how bright a colour is, independent of any hue or saturation that can be assigned. Formally, suppose we have two SPDs: $s(\lambda)$, and $ks(\lambda)$, where $k > 1$. Such a situation might arise with a dimmable light bulb, which emits SPDs all with the same shape, but with different intensities. Then $ks(\lambda)$ will appear as a brighter version of $s(\lambda)$, but otherwise unchanged. In $\mathcal{V}$, $ks(\lambda)$ falls on the same ray as $s(\lambda)$, but farther away from the origin. In fact, all scalar multiples of $s(\lambda)$ fall along one ray, called a *chromaticity ray*, and any two colours on the same ray can always be written, using the correct SPDs, as scalar multiples of each other. The CIE coordinate Y was deliberately chosen to capture luminance information, so a colour with a greater Y-coordinate always appears brighter than one with a lesser Y-coordinate. In the cases of interest in this paper, however, we will usually be comparing two colours on the same ray, so the one farther from the origin has greater luminance, which will be a more convenient geometric criterion.

While luminance encompasses the brightness of a colour, chromaticity expresses all the colour's other aspects, which are often described as hue and saturation. Instead of dealing with hue and saturation directly, chromaticity coordinates are two quantities x and y with no easy perceptual interpretation:

$$x = \frac{X}{X+Y+Z}, \tag{10}$$

$$y = \frac{Y}{X+Y+Z}. \tag{11}$$

All the colours along a ray have the same chromaticity (hence the term *chromaticity ray*). The plane $X+Y+Z = 1$ intersects all the rays that correspond to real colours, so we can project each ray along itself to a single point on the plane $X+Y+Z = 1$. The rays corresponding to real colours lead to a set of points in that plane, which, after some transformations, become the familiar *chromaticity diagram* $\mathcal{C}$ shown in Figure 3.

The boundary of the chromaticity diagram is called the *spectrum locus*, and it is the projection onto $X+Y+Z = 1$ of the 3d spectrum locus. Apart from the straight line segment at the bottom, the spectrum locus contains a point for each monochromatic wavelength (and line segments joining those points if a discretization is used), and only monochromatic SPDs occur on the spectrum locus. The chromaticity diagram also shows empirically that no monochromatic SPD can be written as a sum of any combination of monochromatic SPDs of other wavelengths, so the only way to match the chromaticity of a monochromatic SPD is to use that SPD. If the set of all colours is seen as an irregular convex cone in $\mathcal{V}$, then the chromaticity diagram can be seen as the profile of that cone.

3 Statement of Problem

Even if two SPDs $s_1(\lambda)$ and $s_2(\lambda)$ are metameric, so that T assigns them to the identical colour vector (i.e. $T(s_1) = T(s_2)$), one can have significantly more or less power than the other, because the cones in the human eye respond more strongly to some wavelengths than others. An important practical problem, especially for mobile devices with limited battery life, is to produce a desired colour with as little power as possible. In our terms, given a target colour, we want to select an SPD in that colour's equivalence class that uses the least power, in other words, since all the class's members are metamers, to find a minimal-power metamer.

Some analysis will make this minimal-power metamer problem more tractable. To begin with, suppose that some SPD $s(\lambda)$ produces a colour with vector v. Then the SPD $2s(\lambda)$ produces a colour of the same chromaticity as v, but with twice the luminance. If $s_{\text{min}}(\lambda)$ is a minimal-power metamer, say with power level p_{min}, for v, then $2s_{\text{min}}(\lambda)$ is a minimal-power metamer for $2v$, with power level $2p_{\text{min}}$, so solving the problem of interest for one colour solves it for every colour with the same chromaticity.

Suppose further that we have available only one unit of power, and would like to produce a colour of the chromaticity of v, but with as much luminance as possible. Then the best we can do is use the SPD

$$s_{\text{max}}(\lambda) = \frac{1}{p_{\text{min}}} s_{\text{min}}(\lambda), \tag{12}$$

to produce the colour

$$v_{\text{max}} = \frac{1}{p_{\text{min}}} v, \tag{13}$$

of maximum luminance, and with the same chromaticity as v. We can rearrange Equation (13) to get

$$p_{\text{min}} = \frac{v}{v_{\text{max}}}, \tag{14}$$

so that the "ratio" of the two vectors v and v_{max} can be used to calculate p_{min}.

The proof of this statement and its converse follow by way of contradiction. Suppose that some SPD $t(\lambda)$ produced the colour v_{max} in Expression (13), but with less than unit power. Then $p_{\text{min}} t(\lambda)$, whose power is less than p_{min}, would produce the colour v, contradicting our assertion that p_{min} is the least power needed to produce v. Conversely, assume that v_{max} is the colour of highest luminance, of the same chromaticity as v, that can be produced with unit power, and suppose (by way of contradiction) that the "ratio," given by the right side of Expression (14), is not the minimum power needed to produce v. Then the true minimum power $p_{\text{true}} < v/v_{\text{max}}$, which implies that $(1/p_{\text{true}})v$ has higher luminance than v_{max}, but uses less than unit power, which contradicts the assumed maximality of v_{max}.

To summarize, we have two formulations of the minimal-power metamer problem:

Formulation 1: *Find the SPD of minimal power that produces a target colour v.*
Formulation 2: *Find the SPD of unit power or less that produces a maximum-luminance colour of the same chromaticity as v.*

The previous discussion shows that these formulations are equivalent (in fact, we will see later that they are dual linear programming (LP) problems), so that the solution of either of them immediately follows from the solution of the other. Indeed, Expression (12) says that the two solution SPDs are scalar multiples of each other.

While we started the paper with Formulation 1, we will solve the minimal-power metamer problem using Formulation 2 instead. The main advantage of Formulation 2 is that it leads to easily visualized geometric constructions in the three-dimensional space $\mathcal{V}$, while the first formulation requires working in the 31-dimensional (or higher) space $\mathcal{S}$. A second benefit is that Formulation 2 makes visually explicit a smooth transition from discrete SPDs to continuous SPDs.

4 The Discrete Case

4.1 The Unit-Power Hull

Formulation 2 works in the set of SPDs of unit power or less, so we will construct a geometric object, the *unit-power hull* $\mathcal{U}$, which consists of all colours that can be produced with unit power or less. $\mathcal{U}$ is a subset of three-dimensional colour space V, and can be easily drawn and visualized. In the discrete case, it is a convex polyhedron, which makes it easy to work with. An algebraic development will first be given for $\mathcal{U}$, which will then be translated to a geometric development.

Algebraically, the power P of an SPD $s(\lambda)$ is a linear functional on the vector space $\mathcal{S}$ of all SPDs. For convenience, this presentation will discretize SPDs into 31 channels, going from 400 to 700 nm in increments of 10 nm; the arguments follow through easily, however, for discretizations of any size. The 31 unit-power, monochromatic SPDs $\mathbf{1}_{400}$, $\mathbf{1}_{410}$, $\mathbf{1}_{420}$, ... $\mathbf{1}_{700}$ will serve as a basis for $\mathcal{S}$. Any SPD s can be written algebraically as a linear combination of these basis vectors:

$$s = s(400)\mathbf{1}_{400} + s(410)\mathbf{1}_{410} + s(420)\mathbf{1}_{420} + ... + s(700)\mathbf{1}_{700}. \tag{15}$$

The values of s at the various wavelength channels can therefore be interpreted as coefficients in this linear combination.

Each value of s at some wavelength can also be thought of as the physical power for that wavelength channel, with units such as Watts. For instance, an SPD s with $s(510) = 3$ would contain 3 Watts in the 10 nm wide channel centered on 510 nm. The total power P of s is given by

$$P(s) = s(400) + s(410) + s(420) + ... + s(700). \tag{16}$$

For any real SPD, each coefficient must be non-negative. If in addition the total power is no greater than 1 (of some unit of power), then each coefficient must also be no greater than 1. Even more, the coefficients must sum to no more than 1. Two conditions are therefore required:

$$1.\ \text{For every } \lambda, s(\lambda) \geq 0, \tag{17}$$

$$2.\ \sum_{\lambda=400 \text{ nm}}^{700 \text{ nm}} s(\lambda) \leq 1. \tag{18}$$

Applying T to the vector space $\mathcal{S}$ of SPDs maps the unit-power monochromatic SPDs to the spectrum locus vectors in the three-dimensional vector space $\mathcal{V}$ of colours:

$$T(s) = s(400)T(\mathbf{1}_{400}) + s(410)T(\mathbf{1}_{410}) + s(420)T(\mathbf{1}_{420}) + ... + s(700)T(\mathbf{1}_{700}). \tag{19}$$

In colour space, then, $T(s)$ is a colour that is written as a linear combination of the spectrum locus vectors. Note that while the 31 unit-power monochromatic SPDs are linearly independent in $\mathcal{S}$, the 31 spectrum locus vectors in $\mathcal{V}$ show many linear dependencies. s and $T(s)$, however, share a common set of coefficients, so Conditions (17) and (18) apply equally well in either $\mathcal{S}$ or $\mathcal{V}$.

These two conditions can be interpreted geometrically in terms of convexity. A subset $\mathcal{A}$ of a vector space is said to be *convex*[4] if, whenever two points v_1 and v_2 are in $\mathcal{A}$, every point in the line segment joining v_1 and v_2 is also in $\mathcal{A}$. Algebraically, that line segment is the set of all linear combinations $\alpha_1 v_1 + \alpha_2 v_2$ of v_1 and v_2, where α_1 and α_2 are non-negative and sum to 1. Any linear combination whose coefficients satisfy these conditions is called a *convex combination*. The set of all convex combinations is called the *convex hull* of v_1 and v_2.

The definition extends easily to any finite set of vectors. The convex hull of three non-collinear vectors, for instance, is a triangle, while the convex hull of four non-coplanar vectors is a tetrahedron. If one requires that all coefficients are non-negative, but their sum is less than or equal to 1, instead of strictly equal to 1, then any line segment between the origin and any point in the convex hull is also included. Equivalently, one could simply add the origin as another vector, and just use the convex hull of the augmented set.

This discussion shows that T maps the set of all SPDs that satisfy Conditions (17) and (18) to the convex hull of the 3d spectrum locus (which, we recall, includes the origin as a limit point). Any colour produced as a linear combination of the spectrum locus vectors, whose coefficients are non-negative and whose sum is less than 1, can similarly be mapped back to an SPD in $\mathcal{S}$, just by applying those coefficients to the basis of 31 unit-power monochromatic SPDs, giving an SPD whose total power is less than 1. The SPDs of total power less than 1 are therefore equivalent to the convex hull of the 3d spectrum locus, and we will call this convex hull the unit-power hull, and denote it by $\mathcal{U}$.

Figure 2 shows $\mathcal{U}$ as an explicit subset of three-dimensional colour space $\mathcal{V}$, using the standard 1931 CIE coordinate system for $\mathcal{V}$. The spectrum locus vectors are labeled with their wavelengths. 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.

A striking qualitative feature of $\mathcal{U}$ is the sharp peak that occurs at the spectrum locus vector for 450 nm. Another feature is that most of the vectors corresponding to the longer wavelengths are nearly in the plane $Z = 0$. In combination, these two features mean that $\mathcal{U}$ is shaped somewhat like an irregular cone, with edges radiating from the apex at 450 nm. That radiating pattern will reappear later, in the context of the chromaticity diagram.

Standard convexity theory (Sect. 20 of Ref. 4) tells us that $\mathcal{U}$ is a polyhedron, every vertex of which is the tip of the spectrum locus vector for some wavelength λ. Another convexity theory result, Carathéodory's Theorem (Thm. 2.23 of Ref. 4), says that, given any point u in $\mathcal{U}$, there exists at least one expression for u as a linear combination of no more than four generating vertices, i.e. u can be written as a combination of spectrum locus vectors, where all but four of the coefficients are 0.

A similar, but stronger, result applies occurs for those boundary points that lie on faces that do not touch the origin. We see that every bounding face of $\mathcal{U}$ is a triangle. Suppose that we have a boundary point b inside some triangular bounding face τ that does not touch the origin. Then the *only* expression for u is via barycentric coordinates (Thm. 2.25 & 2.26 of Ref. 4) involving the three vertices of τ. The vertices are the spectrum locus vectors for some wavelengths, so the only SPD, of unit power or less, for u is 0 except at those three wavelengths. Since barycentric coordinates always sum to 1, however, that SPD must in fact require the maximum power of one unit. (The summing to 1 only holds when the origin is not a vertex of the triangle; otherwise some power might be placed at the origin, just outside the visible spectrum, and so not appear in the SPD.) This maximizing result will be used later when investigating minimal-power metamers.

(Further inferences can be made for special cases in which b is on an edge, or even on a vertex. If b is on an edge, then only two wavelengths have any power. If b is actually on a vertex, then only one wavelength has any power, so the only possible SPD is of unit power and monochromatic, so b is actually at the tip of a spectrum locus vector. The fact that *every* spectrum locus vector occurs as a vertex of $\mathcal{U}$ is an empirical finding, and not a mathematical necessity; one implication is that no single-wavelength SPD can be matched by a combination of SPDs of other wavelengths.)

An important point to bear in mind is that while every colour in $\mathcal{U}$ *can* be produced by an SPD of unit power or less (and no colour outside $\mathcal{U}$ *can* be produced with unit power or less), it does not follow that every colour in $\mathcal{U}$ *must* be produced in such a way. In fact, since the cones in the human visual system have weak responses to some wavelengths and strong responses to other wavelengths, one can construct SPDs that uses a considerable amount of power, without provoking much visual response. In other words, a colour can be produced efficiently or inefficiently, and two SPDs of very different physical powers can produce an identical colour. All the metamers that produce a target colour are mapped by T to that colour, but those metamers vary greatly in their physical power requirements, and we would like to find the metamer that requires the least power.

4.2 Minimal-Power Metamers

This paper aims to find a construction for the SPDs of a minimal-power metamer, one that produces a target colour with as little power as possible. In an earlier section, we showed that this problem can be reformulated as finding the maximum luminance with which a colour of a given chromaticity can be produced, using one unit of power. In this section, the unit-power hull $\mathcal{U}$ will be used in a geometric solution of the reformulated problem.

Suppose then that we have a target chromaticity and one unit of power with which to maximize the luminance of that chromaticity. Any colours of the target chromaticity must lie along a ray $\mathcal{R}$ emanating from the origin. To maximize luminance, the unit-power SPD should produce a colour as far as possible from the origin on $\mathcal{R}$. Any colour produced with unit power, however, must lie in $\mathcal{U}$, so the maximal-luminance colour is in $\mathcal{R} \cap \mathcal{U}$, and the farthest such point along $\mathcal{R}$ occurs at the point p where $\mathcal{R}$ exits $\mathcal{U}$. p must therefore be in a triangular face τ of $\mathcal{U}$. Furthermore, τ cannot contain the origin (if it did, then $\mathcal{R}$, which also contains the origin, would have to be a subset of τ, in which case $\mathcal{R}$ would exit through the far edge of τ rather than through the face of τ.)

The preceding geometric construction has solved the reformulated problem by finding the colour p of maximum luminance for a given chromaticity, and we can now move easily to the original problem of finding a minimal-power metamer. As shown in the previous section, p results from a unique SPD σ of unit power which has zero power at all but three wavelengths. σ must also be a minimal-power metamer for p. (If it were not, then a minimal-power metamer σ' would exist, with $P(\sigma') < 1$, and the SPD $\sigma'/P(\sigma')$ would have unit power and give the same chromaticity as p, but at a higher luminance, at the point $p/P(\sigma')$. That point, though, is outside $\mathcal{U}$, which is a contradiction.) Furthermore, since barycentric coordinates are unique, σ is not just *a* minimal-power metamer for p, but *the* minimal-power metamer, so we have also proved a corollary that minimal-power metamers are unique.

The minimal-power metamer for any colour of the same chromaticity as p can now be found simply by multiplying σ by an appropriate scalar factor. Given an arbitrary target colour v, we can apply the constructions just outlined to the chromaticity (x_v, y_v) of v, to find the unit-power SPD σ_v for the colour of maximum luminance for that chromaticity. The appropriate scalar multiple of σ_v is then the minimal-power metamer for v, as required by the first formulation.

Apart from the easily visualized, three-dimensional body $\mathcal{U}$, the geometric approach leads to a convenient interpretation in terms of the chromaticity diagram. We have seen that a colour's chromaticity is needed to determine the shape of the minimal-power SPD, and the luminance is only needed to determine the SPD's scale, so problem solutions should be expressible using just the two-dimensional chromaticity diagram. We will now accomplish that task.

To begin with, we can project $\mathcal{U}$ onto the chromaticity diagram using the transformation in Equations (10) and (11). Geometrically, every chromaticity ray $\mathcal{R}$ (for a real colour) intersects the plane $X + Y + Z = 1$ at some point r, and this transformation collapses every colour of that chromaticity onto r. The curved portion of the boundary of the chromaticity diagram is called the *spectrum locus*, and is the two-dimensional projection of the 3d spectrum locus. It is convex—in fact, strictly convex, in the sense that no wavelength vertex is in the convex hull of the other boundary points.

Under the chromaticity transformation, the triangular face of $\mathcal{U}$, with vertices λ_1, λ_2, and λ_3, goes to the triangle whose vertices are the images of λ_1, λ_2, and λ_3. $\mathcal{U}$ has many triangular faces, but we are only interested in ones through which chromaticity rays can exit, which we have already seen cannot contain the origin. Figure 3 shows the images of those faces in the chromaticity diagram. Note the radiating pattern that seems to occur, with many triangle edges emanating from 450 nm or nearby wavelengths. This radiation corresponds to the feature mentioned earlier, in which the vector at 450 nm formed a sharp apex for $\mathcal{U}$, from which many edges flowed down.

It was seen earlier that (apart from cases where p is on an edge or vertex) a minimal-power SPD for a given chromaticity requires power at exactly three unique wavelengths, corresponding to the vertices of a bounding triangular face of $\mathcal{U}$. As an example, consider the case where the chromaticity ray $\mathcal{R}$ exits $\mathcal{U}$ at the point p, through the face whose vertices are the spectrum locus vectors for 450, 560, and 570 nm. Figure 2 highlights this face in three dimensions. The chromaticity transformation sends this face to another triangle, in two dimensions, which Figure 3 highlights. The minimal-power SPD for any colour whose chromaticity is inside the highlighted triangle will require power only at the wavelengths 450, 560, and 570 nm. The relative power required at those wavelengths will depend on the barycentric coordinates of that chromaticity in the triangle, and the scaling of the minimal-power SPD will depend on the colour's luminance. Every chromaticity in the diagram fits into some unique triangle, and so has a unique set of three wavelengths. Furthermore, most chromaticities require some power at about 450 nm, so this wavelength is more important than the rest, at least for theoretical efficiency considerations.

4.3 The Linear Programming (LP) Interpretation

The first formulation of the minimal-power metamer problem can be naturally expressed in terms of linear programming. A standard LP problem involves minimizing or maximizing a linear functional in the positive orthant of a vector space, subject to various linear constraints about which points of the vector space are feasible solutions. In the minimal-power metamer problem, the vector space is the set $\mathcal{S}$ of SPDs, and the quantity to be minimized is the total power, given by the linear functional

$$P(s) = \sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda). \tag{20}$$

We must produce a target colour, with CIE coordinates $(X_{\text{tar}}, Y_{\text{tar}}, Z_{\text{tar}})$, so we require the following linear constraints on any SPD $s(\lambda)$ in the vector space $\mathcal{S}$:

$$\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{x}(\lambda) = X_{\text{tar}}, \tag{21}$$

$$\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{y}(\lambda) = Y_{\text{tar}}, \tag{22}$$

$$\sum_{\lambda=400\text{ nm}}^{700\text{ nm}} s(\lambda)\bar{z}(\lambda) = Z_{\text{tar}}. \tag{23}$$

This LP problem is simple to implement using any LP solver, and is numerically undemanding, at least for a 31 channel discretization. Furthermore, standard LP theory (Theorem 1.10 of Ref. 5) says that, since there are three constraints, at most three entries of any solution vector can be non-zero. In our case, a solution vector consists of the 31 values for $s(\lambda)$, as λ varies over the wavelengths from 400 to 700 nm, so a minimal-power $s(\lambda)$ has positive power at no more than three wavelengths. We have already derived this result geometrically, using the triangular faces of $\mathcal{U}$.

Any LP problem in standard form leads to an equivalent problem, called the dual LP problem (p. 173 of Ref. 4), obtained by switching the functional and the constraints. In this paper, the second formulation is in fact the LP dual of the first formulation. For clarity, we proved the equivalence of the two formulations directly, but it could also have been derived by algebraic manipulation of Equations (20) through (23).

While an LP problem and its dual are mathematically equivalent, often one formulation is more easily interpreted than the other, and provides an easily navigated geometric picture. The first problem cannot be visualized because it is situated in a 31-dimensional space. While each solution provides a set of three unique wavelengths, those three wavelengths cannot be viewed as triangles, as we did so naturally, and there is no easy way to see that the shape of the minimal SPDs depends on chromaticity rather than luminance, so a simple presentation like Figure 3 would never have been reached. The importance of the wavelengths near 450 nm is also obvious from that figure, but is hidden in the LP setting. Furthermore, we will see in the next section that the second formulation can easily be extended to the case of continuous SPDs, while it would be difficult to analyze the behavior of Equations (20) through (23) as the discretization becomes increasingly fine.

On the other hand, the first formulation is very useful for calculations. Equations (20) through (23) can be entered in an LP solver in a few minutes, and any computer can calculate a target colour's minimal SPD, encompassing both shape and scale, in less than a second. While the second formulation makes it easy to read off the three non-zero wavelengths, calculating barycentric coordinates can be an involved calculation, as can determining the correct scale factor, and writing a computer program to accomplish these tasks could be time-consuming.

Given the complementary advantages of the two approaches, the best results would most likely be obtained by using the first formulation for calculations, and the second formulation for analysis.

5 The Continuous Case

The derivations so far have treated SPDs as discrete. A 10 nm increment was used consistently, although the constructions and arguments follow through easily for finer increments of 5 nm, 2, 1, 0.1, etc. In the limit, the increment goes to 0, and discrete SPDs become continuous SPDs. This section will show that continuous minimal-power metamers are non-zero at only two wavelengths, instead of the three that are typically required for discrete metamers.

This result can be easily seen in the figures. Suppose that we have maximized the luminance for a colour in the discrete case, by finding that its chromaticity ray exits $\mathcal{U}$ at the point p marked in Figure 2. p is contained in the triangular face bounded by 450, 560, and 570 nm, which the figure highlights. Figure 3 highlights the same point and triangle, but in the chromaticity diagram. These calculations used a discrete increment of 10 nm.

Now shrink that increment, say to 2 nm. Then the bounding spectrum locus will be subdivided more finely; four new vertices will be added between 560 and 570 nm, at 562, 564, 566, and 568 nm. Also, new vertices will appear around 450 nm, at 448 and 452 nm. There will be a new triangular face containing p, whose vertices are multiples of 2 nm. The vertex at the apex might move along the boundary a bit, to the left or the right, or it could just as easily stay at 450 nm. The base of the new triangle, however, is now only 2nm wide, rather than 10. The new triangle as a whole is narrower and closes in on p more. As the increment shrinks further still, to 1 nm, or 0.1 nm, each new triangle is narrower still: its sides approach p more, and its base is progressively smaller. In the limit as the increment goes to 0, the triangles collapse to a line segment which contains p. The power in the two wavelengths in the base gets concentrated in a single wavelength somewhere between them, while the power in the wavelength at the apex stays about the same. The resulting continuous SPD therefore has power in only two wavelengths, as was to be shown.

(The foregoing proof depends implicitly on the fact that the various wavelengths appear in order on the spectrum locus, so that the vertices for 562, 564, 566, and 568 nm occur sequentially between 560 and 570 nm in a geometric sense, rather than just in a numerical sense. An inspection of Figure 3 shows, however, that every triangle can be considered to have an obvious apex and base, and that the vertices of the base are two *consecutive* wavelengths. The assertion that the triangles converge to a line segment as the increment goes to 0 relies on the fact that the vertices of the new base are between the vertices of the old base, and thus approach each other ever more closely. The consecutive ordering of the wavelengths on the spectrum locus is an empirical finding rather than a mathematical conclusion, so it must be observed rather than derived.)

6 Relation to Previous Work

This paper's main contribution is a unified geometric derivation of results about minimal-power metamers. Many of the results themselves have been known for some time, although their origins are sometimes obscure, especially in the continuous case.

In a 1950 note[6] published in JOSA, MacAdam asserts that a maximum-luminance SPD "can be obtained in only one way, by mixture of suitable intensities of a single pair of wavelengths." He gives no evidence for this assertion, however. An earlier paper[7] of his does contain some computations involving colours that are complementary in the sense that their chromaticities are collinear, in the chromaticity diagram, with a neutral white (MacAdam used Illuminant C). If such colours were chosen to be on the spectrum locus, then an SPD for Illuminant C would result, with power at only two wavelengths. The pair of two wavelengths that produced Illuminant C to a desired luminance, but with minimal power, could then be found by an exhaustive check (MacAdam found the pair 448 and 568.7 nm). SPDs that are non-zero at exactly three wavelengths are also mentioned as possibilities, so checking some of those could reveal that an SPD with two non-zero wavelengths is the most efficient. Still, it seems implausible that MacAdam jumped from the single case of Illuminant C to a general assertion about all chromaticities. Perhaps an elegant insight that seemed too obvious to mention led to his conclusion, but the reasoning is still unknown.

The author has failed to find an earlier explicit statement of the LP approach to the discrete minimal-power metamer problem, but this approach seems rather routine, especially when one considers that LP methods were used in colour science at least as early as 1972,[8] for both spectral and metameric matching. Likely the only original contribution of the current paper to the LP approach is the recognition that the dual problem's geometric interpretation makes it more instructive than the primal problem.

7 Summary

This paper investigated the problem of finding the minimal-power metamer for a target colour. The main novel contribution was a geometric object called the unit-power hull, which was constructed from some well-established mathematical objects in colorimetry. The unit-power hull allows a natural, easily visualized solution to a problem that is equivalent to the original: finding the colour of maximum luminance but fixed chromaticity that can be produced with a set amount of power. Furthermore, this geometric approach extends readily from metamers produced by discrete SPDs to metamers produced by continuous SPDs, and shows why their minima have different forms. It is hoped that the concrete geometric development presented here will lead to further understanding for other optimization problems in colour science.

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. James A. Worthey, "Vectorial Color," Color Research & Application, Vol. 37, Number 6, December 2012, pp. 394-409.
4. S. R. Lay, *Convex Sets and Their Applications*, Dover Publications, 2007.
5. Bernard Kolman & Robert E. Beck, *Elementary Linear Programming with Applications, 2nd ed.*, Academic Press, 1995.
6. David L. MacAdam. "Maximum Attainable Luminous Efficiency of Various Chromaticities," JOSA, Vol. 40, Issue 2, 1950, p. 120.
7. David L. MacAdam. "Photometric Relationships Between Complementary Colors," JOSA, Vol. 28, April 1938, pp. 103-110.
8. Noboru Ohta & H. Urabe, "Spectral Color Matching by Means of Minimax Approximation," Applied Optics, Vol. 11, Issue 11, November 1972, pp. 2551-2553.

