# Linear Models of Reflective Colour

by

Alan William Paeth

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 1993

I hereby declare that I am the sole author of this thesis.

I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

# Abstract

Colour is an artifact of the continuous, interacting physical processes of light and matter. In theory, colour models using linear systems require a high dimension. In practice, the perception of light rays is well-modeled using an 3-space whose construction is based upon the trichromatic nature of human vision. Linear systems may also model the reflectance functions of surfaces (Yilmaz, 1962). A larger $n$-space is required to minimize perceivable colour error of a surface under unspecified illumination. Bounds as low as $n \approx 6$ have been demonstrated by using a Fourier basis (Buchsbaum and Gottshalk, 1984). Contemporary research in perceptual models for human and machine vision apply regression techniques, such as the singular value decomposition, to create orthogonal bases (Maloney, 1986), often derived from widely available reflectance data on natural surfaces (Krinov, 1947).

This thesis extends methods of low-dimensional reflectance modeling, thereby providing a framework suitable to computer graphics with direct application to radiosity-based rendering and digital printing. The framework systematizes both the sequential production of orthogonal bases and methods of dimensional change. Each basis spans a *reflectance space,* by analogy to the additive colour spaces of light. A reflectance space contains a unique *reflectance solid:* a convex polytope whose volume describes merely valid colours (e.g., excludes surfaces of "negative" reflectivity). The SVD methods are constrained so that white reflectances $\rho(\lambda) = 1$ are always representable. This concedes little in accuracy yet is both necessary and sufficient to guarentee that all reflectance solids exhibit a useful symmetry of inversion. Solids are derived from spectroradiometric data produced in-house and studied using both formal and empirical methods. Digital production of colour facsimile whose reflectances form an equivalence class with an original's engenders an illumination-independent means of colour correction, named *reflectance matching.*

(Dedication to be added during final binding)

## PREFACE

To paraphrase an ancient riddle,

> "A redwood falls during the dark of night. What is its colour?"

The old adage offers the answer "none: it is unseen". A colour scientist familiar with the CIE-defined standard observer might venture that "absence of a source of illumination implies no chromaticity coordinates: the colour is undefined". The computer graphics specialist suggests the $RGB$ triple $[0, 0, 0]$ and considers this descriptor "black".

In fact, the tree bark is *red* and so remains, unperturbed by its fall, the darkness of night or the absense of any observer. Its colour is an intrinsic property of its surface, not its setting.

This thesis explores linear models of colour representation suitable for digital computation which are consistent with this axiomatic physical-surface model.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Broadly, this thesis studies linear models of reflective colour[1] – the colour which surfaces take on in the presence of a source of illumination, yet which are nonetheless an intrinsic property of the surface without reference to any source. Whereas a traditional *additive* colour model describes the mixing of light and is linear[2], the *subtractive* model used to describe the loss of energy when a ray reflects off surfaces is a multiplicative phenomena and is not linear. Linear models are widely used throughout mathematics and many places within computer graphics. Linear reflectance models are a desirable extension: misapplications of additive formulae in problems of a subtractive nature remain common. The choice of linear reflectance models parallels research in machine vision, notably *colour constancy* (Sällström, 1973), (Buchsbaum, 1980), (Shafer, 1985), (Young, 1986), (Gershon, 1987), (Tominaga and Wandell, 1989), (Tominaga and Wandell, 1990).

Reflective colour describes commonplace physical phenomena. Thus, digital

[1]Canadian spelling is chosen based on an 1890 order-in-council (Gage, 1985).

[2]Following pioneering work in colorimetry and algebra of the last century (Grassman, 1853).

colour systems seeking realism in advanced rendering, colour reprographics or other pertinent research areas must employ this model. Given perfect knowledge of a surface's *reflectance function,* its physical behaviour in the presence of light rays can be precisely described. As this is impossible in practice, a useful characterization which approximates this behaviour must degrade gracefully in providing an accurate description of a surface within a computationally tractable framework.

Throughout this thesis, surface reflectances are represented using linear combinations of model elements. The basis elements represent fictitious (and very often, impossible) reflectances; the summing of reflectances is an abstraction having no physical analog. Nonetheless, the methods make possible a concise descriptions of surface reflection, a provide a computational framework making simple the phrasing and solution of related problems. Regression techniques and extensions thereto are used to produce suitable sets of basis elements. These span a linear *colour space* in which perceptual authenticity of reflective media is maintained using models having reduced dimension. In addition, methods of *reflectance matching* are described and explored. By this technique a facsimile is produced having a reflectance function which is equivalent to an original's within the common subspace. The match exists without reference to an external source of illumination and represents a new and original approach to the solution of difficult and ongoing problems in reprographic colour correction.

Linear algebra simultaneously supports the examination of linear colour space using Euclidean geometry. Formally, the basis elements span a linear *reflectance space;* points within the space are associated with (reflectance) functions. Considering merely the locus of points whose associated reflectances are physically possible reveals the *reflectance solid* unique to that space. Similarly, the set of points describing the range of reflectances which can be created on a specific physical device, such

as a digitally-controlled printer, define its *reflectance gamut,* by analogy to similar models common within the additive domain.

The geometric study of the reflectance solids comprises an essential core of the thesis. Modifications to the basis elements to better account for colorimetric properties of typical reflectances imparts a high degree of symmetry to the solids. This gives further insight into the nature of reflectance space. This substantial formalization allows exploration of a diverse set empirical problems. All employ a *spectroradiometer* (which quantizes and records the reflectance function of a surface at high precision) and a thermal sublimation printer, both under direct digital control.

## 1.1 Overview

The thesis is arranged as eight chapters and eight appendices. The chapters describe the evolution of a computational framework derived and verified using both formal and empirical techniques. The first three chapters describe the phenomena and basic models of colour in terms readily extended beyond the additive domain. Chapters four and five formalize the models in the reflective domain in terms both of linear algebra and of an accompanying constructive geometry. Two related chapters (6 and 7) apply the newly established models to a range of problems concerning empirically-sampled media. The final chapter discusses present and future thesis extensions. Each chapter is outlined in the paragraphs below.

Chapter 2 presents the fundamental operations involving rays and surfaces and describes them using a consistent linear algebra. Based on this, the *colour space* is formally defined. The general model can accommodate conventional models of both additive and subtractive colour, arranged as a taxonomy called the *colour hierarchy*

appearing as Figure 2.1. A parent $n$-dimensional model based on an orthogonal (diagonal) set of functions lies at the root of the hierarchy and becomes the underlying representation common to all linear models. (Throughout the thesis, $n$ describes this dimension, most often defined by the resolution of the spectroradiometer used to create any data). This model, described in connection with Figure 2.2, formalizes many of the first-principle approaches in common use.

Chapter 3 examines colour spaces in the context of light rays, thus formalizing *additive colour spaces.* Although the domain is well understood, formalization using linear algebra remains a subject of ongoing study (Shafer, 1982), (Marimont and Wandell, 1992). For present purposes, the formalization gives rise to the conditions of *metamerism,* whereby two sources of illumination are described by identical *illumination coordinates* in a space of reduced dimension $m$. The derivation makes extensive use of matrix algebra and the concepts of generation and detection as defined in Chapter 2. In the second section, the geometry of the space is studied with particular regard to those the convex solids whose locus of points describe physical sources. In the process, traditional forms such as the chromaticity diagram are rederived in the language and context of the thesis. More important, the concept of the *illumination solid* is described and its defining linear algebra made familiar.

Chapter 4 examines colour spaces in the context of reflective surfaces, thus formalizing *reflective colour spaces.* This framework is relatively unexplored. Following lines of exposition paralleling the previous chapter, *reflectance matching* is described. Its related linear algebra need employ merely one matrix (the reflectance basis) or its transpose to map between *reflectance coordinates,* which are the $m$ components of any point lying in reflectance space, and *reflectance vectors,* which are the $n$ components approximating a continuous reflectance function in discrete form. The second section describes *reflectance solids,* which unlike their

illumination analogs are finite as well as convex. Their geometry is described with particular attention paid to those features which are not artifacts of sampling but which persist in reflectance spaces of increasing dimension. Both empirically determined and canonical solids (in spaces having predetermined orthogonal bases) are constructed and rendered. The disparity of their appearance and the algebraic liabilities thereby uncovered suggest a means of modification in the empirical derivation of basis elements.

Chapter 5 considers methods of modified orthogonal basis construction which guarantee the presence of an initial basis element $\bar{\beta}_1(\lambda) = 1$ which describes an *ideal white* reflectance. Two methods are considered, the second is chosen. In each case the constrained basis set is compared using standard techniques against the standard regression model (which is guaranteed to minimize overall variance in approximating the surface reflectance samples from which it is derived). The study suggests that the representational loss arising from the modified method is both minimal and is quickly won back with dimensional increase, suggesting its universal adoption as a means of basis production and representation. The second half of Chapter 5 considers prototypical reflectance solids all possessing central symmetry. This attribute is a direct consequence of modification, shown formally by algebraic methods. These solids become the *zonotopes:* geometric objects of an important crystal symmetry class (Loeb, 1970) of both historical and present day study.

Chapter 6 describes deflation and reflectance matching. The chapter begins with a description of both the traditional methods of colour correction and the radiometric practice used here to produce reflectance data. Deflation provides a reduced parameter model wherein reflectance matches may occur. The final three sections of the chapter consider deflation and matching based first upon globally determined bases using traditional regression, then upon global bases using the

symmetric spaces of Chapter 5, and lastly upon local methods which employ a multiplicity of bases. The chapter contains the empirical core of the thesis represented in empirical tables and graphs. The final section also demonstrates the manipulation of reflectance data lying on dissimilar bases both by matrix algebra and through pseudocode implementations. (An equivalent production system written in C underlies the creation of most of the artwork and images appearing in the thesis, pseudocode is otherwise used for illustration).

Chapter 7 describes inflation, a near complement to deflation. Because dimensional increase cannot introduce data where none existed before, the study moves toward a geometric analysis of reflectance solids so that the higher dimensional model maintains authenticity of match over the greatest possible range of unspecified sources of illumination. An interactive means of inflation is then described which takes good advantage of the custom (reflectance) solid modeler created in support of the research. The study formalizes the concepts of *opponent colours* extended into the reflectance domain. The chapter concludes with a study of the *reflectance belts* of the symmetric solids, suggesting a means of approximation using convex polyhedra of reduced dimension when the geometric models are studied in higher dimensions.

Chapter 8 presents follow-on work and describes related applications. The work is presented in order of immediacy of extension, beginning with a detailed characterization of transmissive media and concluding with suggestions for further modification. The first section provides a complete characterization of a reflectance gamut at very high sampling density (4096 coloured transmissive surfaces each characterized at 512 wavelengths). The analysis takes advantage of a commercially available software package to perform the regression. The production scripts are included together with detailed plots of the film's gamut. The second section describes

specialized hardware useful in displaying pixel data containing either traditional illumination coordinates or the reflectance coordinates described here. The third section describes methods of extension in the production of orthogonal reflectance elements; a useful non-orthogonal basis is also described. The thesis concludes with a unified description of rendering by describing reflectance-based colorimetric rendering in relation to and in analogy with the methods of spatial rendering which are universal to computer graphics.

Appendices and a detailed bibliography complete the thesis.

# Chapter 2

# Colour and Linearity

This chapter reviews the basic physical models of colour with emphasis on linear algebraic forms. First, the four laws of colour are presented. These describe the physics of rays, surfaces and their interaction, thereby describing colour as it arises in everyday situations. The first three laws are easily represented using a linear algebra which describes the additive methods of coloured lights. The fourth law describes reflection of light by surfaces. This physical process is not well-described by linear equations.

The utility of linear systems together with their well-known role in the modeling of additive colour suggests their universal adoption. In addition, linear systems support a parallel, geometric study and interpretation. This chapter concludes by outlining and formalizing the requirements of a *linear colour space.* As seen in the next two chapters, colour spaces serve as a framework which simplifies the description and interpretation of physical colour. In Chapter 3, traditional methods of (additive) colorimetry are described in the notation and methods described here. Chapters 4 and beyond then appropriate these algebraic and geometric models

to describe colour in a reflective setting. Geometric studies of reflective colour represent new research.

## 2.1 Background

Colour is sometimes ascribed to a ray of light, sometimes to a reflective surface. More precisely, light and matter both possess radiometric properties. The light ray is characterized by the distribution of its constituent electromagnetic radiation; an object's surface is characterized by its action in redistributing the energy of any impinging ray through absorption or reflection. Directly or indirectly, these properties govern the spectral power distribution of a ray of visible light: radiation whose wavelength lies within between roughly 400nm and 700 nm[1] of the electromagnetic spectrum.

When a ray is detected by the eye, a psychophysical response is elicited, ultimately giving rise to a sensation of colour. The perceived colour is governed largely by the distribution of energy within the ray, though other factors such as contrast (Ratliff, 1965), constancy (Wandell, 1989) and assimilation and chromatic adaptation (Stiles, 1961) can come into play. The latter can be significant, as first demonstrated early in this century (von Kries, 1904). Wright's pioneering work on the binocular presentation of perceptually dissimilar sources (Wright, 1934) serves as a common tool in exploring observer adaptation under changing conditions (MacAdam, 1956), and is useful in creating alternate models of colour vision (Land, 1977). Higher-order perceptual models are reviewed in accessible journals (Nassau, 1980).

[1] 1nm = $10^{-9}$ metre.

Disregarding these effects, as when the viewing environment is well-controlled and presented uniformly to both eyes, the colours perceived in ordinary settings are directly determined by the radiometric properties of a light ray. For the purposes of this thesis "colour" will be considered merely in terms of continuous spectral distributions of light and matter. Colour models based entirely upon a standard observer and the direct action of a collection of rays are termed *additive.* Surfaces indirectly take on colour as a consequence of their interaction with light rays. Colour models which predict the consequences of the interaction of light and matter, as by transmission or reflection, are called *subtractive.* As suggested in the introduction, the most advanced subtractive models describe the behaviour of systems of surfaces without reference to any source of illumination.

## 2.2 Fundamental Operations

Central to any system of colour is the light ray – electromagnetic radiation possessing both spatial and radiometric attributes. In rendering models these attributes are often coupled: surface colour may vary as a function of six variables which describe the wavelength(s) of interest, and the bearing and azimuth of the incident and reflected rays common to both source and observer. Simpler models which omit spatial orientation are considered in nearly all the discussion which follows; spatial extensions are described in Chapter 8.

Physical systems of light are based on four fundamental operations. A light ray may be generated by emission from some object, mixed in "additive" fashion with other rays, attenuated by "subtractive" surface reflection and ultimately detected. The first and last operations define unique points of origin and termination for any light ray, the latter often within the eye of a (standard) observer. In contrast,

the intermediate operations of composition carry one or more light rays into one or more new light rays. These allow the systematic creation of rays having new spectral power distributions by the superposition or selective attenuation of any ray's spectral energy.

A viable colour model includes these intermediate operations, for it is by these that new spectral power distributions, and hence new colour, are created. Colour created by the mixing of light is an *additive* system. An example is the reflection seen from a window combined with the scene visible through the window. Colour created by preferential surface reflection is called *subtractive* colour. Examples abound in every day settings; words on the printed page are a common example.

To describe the operations of light the "operands" must first be introduced.

### 2.2.1 Rays and Surfaces

Central to all discussions of colour is the light ray, the ultimate medium by which colour detail is transmitted to the eye. Rays may be generated as seen in equation[2] [2.9], superposed [2.19], reflected [2.21] and detected [2.12]. Rays possess a spectral power distribution $\Phi_\lambda$ which is always non-negative:

$$\forall\lambda: \quad \Phi_\lambda \geq 0. \qquad \text{[illumination law]} \qquad [2.1]$$

The distribution measures the amount of energy present for a parameterized wavelength $\lambda$.

Surfaces possess intrinsic colour, which is imparted to a reflected ray by redistributing the energy of an incident ray. Any surface has a dimensionless reflectance

[2]Equation are presented in the form $[c.n]$ with $c$ the chapter number and $n$ the number of occurrence within the chapter.

function $\rho(\lambda)$ which describes the ratio of reflected flux $\Phi_\lambda^{(\mathrm{r})}$ to incident flux $\Phi_\lambda^{(\mathrm{i})}$ of any ray as a function of wavelength:[3]

$$\rho(\lambda) = \frac{\Phi_\lambda^{(\mathrm{r})}}{\Phi_\lambda^{(\mathrm{i})}}. \qquad [2.2]$$

For surfaces which are neither self-luminescent nor fluorescent (Pringsheim, 1949) $0 \leq \Phi_\lambda^{(\mathrm{r})} \leq \Phi_\lambda^{(\mathrm{i})}$ at any wavelength. This leads to the important physical condition of reflectance:

$$0 \leq \rho(\lambda) \leq 1, \qquad \text{[reflection law]} \qquad [2.3]$$

which is an *isotropic* (orientation independent) model defining a constant reflectance at $\lambda$, as characterizes any *Lambertian* (diffuse reflecting) surface. Extensions which account for the (spectral) bidirection reflectance-distribution function or *(S)BRDF* $\rho(\lambda_{in}, \lambda_{out}, \psi_{in}, \psi_{out}, \theta_{in}, \theta_{out})$ are treated in Chapter 8.

### 2.2.2 Generation

Generation describes a source of light (real or hypothetical) whose continuous spectral power distribution $\Phi_\lambda$ may be regulated by a single argument $\alpha$, with $\alpha \geq 0$. (Since generation is a fundamental operation, a distribution modeled by more than one parameter would be broken down into combinations of this atomic form). That is,

$$\Phi_\lambda^{(\alpha)} = \Phi_\lambda(\alpha), \qquad [2.4]$$

in which the distribution indexed by $\alpha$ is defined as a function of $\alpha$. For instance, the spectral power distribution of a black-body is a function of its absolute temperature,

[3]The parenthetic superscript indicates the physical entity to which the function is related. This usage is borrowed (Wyszecki and Stiles, 1982) and avoids the use of multiple subscripts.

based on *Planck's formula* which is:

$$\Phi_\lambda^{(\mathrm{T})} = 2\pi hc^2 \left[ \frac{\lambda^{-5}}{e^{\frac{hc}{k\mathrm{T}\lambda}} - 1} \right] \quad (W \cdot m^{-3}) \qquad [2.5]$$

with T in degrees Kelvin (°K), $c$ the speed of light, $h$ Planck's constant and $k$ Boltzmann's constant (see SG). According to the *Stefan-Boltzmann Law,* total radiant excitance (hereafter called *intensity)* $M_e$ is proportional to the fourth power of absolute temperature:

$$M_e(T) = (\frac{2\pi^5 k^4}{15h^3c^2})T^4 \qquad [2.6]$$

suggesting expressions of the form

$$\Phi_\lambda^{(\alpha)} = M_e(\alpha) \cdot \phi_\lambda(\alpha), \qquad [2.7]$$

in which $M_e(\alpha)$ records intensity (with suitable physical units) and follows the illumination law: $0 \leq M_e(\alpha)$. The normalized emissivity factor $\phi_\lambda(\alpha)$ represents a dimensionless distribution enclosing a unit area. Though it describes a fictitious source, $\phi_\lambda(\alpha)$ is also everywhere non-negative, and satisfies the stronger condition $0 \leq \phi_\lambda \leq 1$ suggestive of [2.3].

The profile of the black-body source is a broad, continuous distribution having a maximum at a wavelength which increases inversely with temperature. This is the *colour temperature* of the black body, which moves from dull red to blue hot with increasing T. Such sources are valuable to spectroradiometry[4] as their broad distribution maximizes the numerator of [2.2] over a wide range.

The factorization in [2.7] partitions the information governing intensity $M_e$ and hue $\phi_\lambda(\alpha)$. Unfortunately, the latter is parameterized by $\alpha$, indicating a possible

[4]High-temperature ($\approx 3000°C$) Quartz-Iodide Tungsten sources supported all the spectroradiometry of this thesis. Their spectral power distributions closely approximate the distributions of black bodies having colour temperature reduced by $\approx 40°C$.

change in colour temperature as a function of T. Experiments in the mixing of light do well to avoid lamps controlled by rheostats, as both absolute and relative spectral output vary. More useful sources show spectral invariance of the governing parameter. The dependence on the hue function may then be dropped. Throughout the remainder of the thesis this form is employed for reasons justified:

$$\Phi_\lambda = M_e(\alpha) \cdot \phi_\lambda. \qquad [2.8]$$

The hue (emissivity) component can be now considered a basis element which is no longer a function of $\alpha$ but undergoes only a scaling by the intensity factor $M_e(\alpha)$. When $M_e(\alpha)$ is not linear yet monotonic it can be linearized. A general equation for *linear generation* is thereby constructed

$$\Phi_\lambda = G(\alpha) = \alpha\ \phi_\lambda, \qquad \text{[ray generation]} \qquad [2.9]$$

in which $\alpha$ is now a linear scale.

Fluorescent materials demonstrate linearity – their relative spectral power distribution is largely invariant with stimulation (as by ultra-violet light, X-radiation or electron beams) as a consequence of quantum-mechanical effects, while reradiated visible flux is in proportion to incident flux.

The phosphors within a cathode-ray tube are fluorescent and exhibit a high degree of spectral invariance (Cowan and Rowell, 1986). Thus, the cathode ray tube (CRT) is used in this thesis as a canonical device for linear colour generation. Its phosphors are stimulated by a stream of electrons emitted by a gun at the cathode and are accelerated toward the positively charged anode at the tube face. The light intensity is governed by the control voltage V on the CRT's grid, an intermediate electrode which regulates the number of electrons which may reach the phosphors at the anode. Increasing the (positive) grid voltage V reduces the

tube's space charge, allowing more (negative) electrons to pass. At maximum drive $V_0$ the highest light intensity is achieved. A typical equation for the production of light is

$$\Phi_\lambda = k(\frac{V}{V_0})^\gamma \cdot \phi_\lambda^{(red)}, \qquad [2.10]$$

in which control voltage V does not appear as an argument in the red phosphor's spectral power distribution $\Phi_\lambda^{(red)}$. Total light energy grows roughly as a quadratic function of grid voltage, approximated by the term $()^\gamma$ with $2.2 \leq \gamma \leq 2.5$ typical. The system may be easily linearized by compensating drive electronics, or by a look-up table (Catmull, 1979). This process is called "gamma correction" after the exponent appearing in [2.10]. A frame buffer so linearized generates light rays having an invariant distribution yet with total flux under linear control of the associated red pixel value $\mathbf{p}_{red}$ associated with the phosphor spot. That is,

$$\Phi_\lambda = \mathbf{p}_{red} \cdot \phi_\lambda^{(red)} \qquad [2.11]$$

which conforms exactly to the linear generation model stated in [2.9].

### 2.2.3 Detection

Detection is the complementary action to generation. It provides a terminal (and not original) point for any ray of light. More important, detection quantifies the spectral power distribution of a ray, producing a single real value. Thus, it maps a continuous structure onto a discrete one:

$$d = D(\Phi_\lambda). \qquad \text{[ray detection]} \qquad [2.12]$$

Most physical detectors produce an electrical output in proportion to incident photon flux, whether by photoelectric effect (photodiodes) or by resistive change with

temperature (bolometers). The scalar output may be modeled by an integral which sums the response potential across the wavelengths of incident flux.

Detectors are not equally sensitive at all wavelengths as the conversion efficiency is not constant. This condition is accommodated by introducing a dimensionless weighting function $R(\lambda)$, called the responsivity. In algebraic terms,

$$d = D(\Phi_\lambda) = \int \Phi_\lambda R(\lambda) d\lambda. \qquad [2.13]$$

When detectors are not driven into conditions of saturation, the detection is modeled as a linear operation, made clear by the presence of the integral in [2.13]:

$$D(\alpha \cdot \Phi_\lambda) = \alpha \cdot D(\Phi_\lambda) \qquad [2.14]$$

with $\alpha$ as in [2.9]. This leads in particular to

$$D(G(\alpha)) = \alpha \cdot D(G()) \qquad [2.15]$$

indicating that the detected response of a generated source under linear scaling equals the scaled detector response of the element of generation.

Most physical detectors do not show an inhibiting response in the presence of stimulation[5]. This means that $R(\lambda)$ is everywhere non-negative. Normalizing this responsivity puts it in the familiar form $0 \leq R(\lambda) \leq 1$, resembling [2.3].

Abstract detection describes the synthesis of arbitrary responsivities through linear combination of the detector responsivities. Using $D$ as an operator,

$$[D^{(a)} + D^{(b)}](\Phi_\lambda) = D^{(a)}(\Phi_\lambda) + D^{(b)}(\Phi_\lambda) \qquad [2.16]$$

in which the response of an abstract detection operator formed as the sum of two physical operators $D^{(a)}$ and $D^{(b)}$ equals the sum of the individual responses. As

[5] Exceptions include solarization and Sabbiteur effects by which highly overexposed film shows a reversal in optical density.

before, the abstract detector provides a linear response, as it likewise defines coordinate axes in a space of additive colour.

### 2.2.4 Cone Detectors

Most often, the independent detector set is assumed to model the eye, whose retina contains colour sensitive *cones* which fall into three distinct classes for an observer having normal vision[6]. Their responsivities are well modeled by the three *tristimulus vectors* $\bar{x}_2(\lambda)$, $\bar{y}_2(\lambda)$ and $\bar{z}_2(\lambda)$, defined by the Committee International E'clairage (CIE) in 1931 within a 2° field of view for a *standard colorimetric observer* (Commission Internationale D'Clairage, 1933) having normal vision.

The tristimulus functions form an abstract detector set, related to the actual cone responsivities by a linear change of basis. That is, they span the same perceptual space. By design, the tristimulus vectors are everywhere non-negative and the function $\bar{y}(\lambda)$ is the luminous efficiency $V(\lambda)$ (a measure of subjective brightness) previously defined by the CIE in 1924. In canonical form the functions are presented as a table of three columns indexed by row (wavelength) and normalized so that each column sums to unity (that is, each discrete responsivity function encloses an equal area).

An arbitrary $\Phi_\lambda$ produces three tristimulus values $X$, $Y$ and $Z$ which uniquely determine its perceived colour. That is,

$$\begin{aligned} X &= \int \Phi_\lambda \, \bar{x}(\lambda) \, d\lambda \\ Y &= \int \Phi_\lambda \, \bar{y}(\lambda) \, d\lambda \\ Z &= \int \Phi_\lambda \, \bar{z}(\lambda) \, d\lambda \end{aligned} \qquad [2.17]$$

[6]Common forms of colour blindness may be modeled as the absence of one detector class, which reduces the dimensionality of the detector set to two.

are coordinates which determine a unique point in a three-dimensional colour space, traditionally called the *tristimulus space* (Wyszecki and Styles, 1982b). This may be considered a *detection space* of low dimensionality (three) in which perfect predictions of the perceptual matching of colour lights may nonetheless be conducted.

### 2.2.5 Ray Superposition (Additive Mixing)

Light rays may be superposed to form a new ray. This is traditionally called the "mixing of light". Notationally, a light ray whose colour $C$ is composed of two rays having colours $C_a$ and $C_b$ is written

$$C = C_a \oplus C_b, \qquad [2.18]$$

with $\oplus$ the operator of additive composition. Superposition is a natural physical phenomena of light: rays superimpose freely without interference (save for mixing of coherent, monochromatic sources, as with lasers). Conservation of energy indicates that the spectral power distribution $\Phi_\lambda^{(\mathrm{C_a} \oplus \mathrm{C_b})}$ of the ray having colour $C_a \oplus C_b$ is formed by tallying the energy from the distributions present in all constituent rays:

$$\Phi_\lambda^{(\mathrm{C_a} \oplus \mathrm{C_b} \oplus \cdots \oplus \mathrm{C_z})} = \Phi_\lambda^{(\mathrm{C_a})} + \Phi_\lambda^{(\mathrm{C_b})} + \cdots + \Phi_\lambda^{(\mathrm{C_z})}. \qquad \text{[ray mixing]} \qquad [2.19]$$

That is, the spectral power distribution of a composite ray is formed by summing the power distributions of its constituent parts. Composition of coloured light under operator "$\oplus$" takes the form of addition between the terms on the right-hand side of [2.19], hence the name "additive colour". It is understood that the constituent terms are summed along matching wavelengths, as the subscript $\lambda$ indexes the spectral energy of each source.

The response of a sum of sources $a$ and $b$ having spectral power distributions $\Phi_\lambda^{(\mathrm{a})}$ and $\Phi_\lambda^{(\mathrm{b})}$ may once again take advantage of the detector linearity shown in

[2.14]. Specifically,

$$D(\Phi_\lambda^{(a)} + \Phi_\lambda^{(b)}) = D(\Phi_\lambda^{(a)}) + D(\Phi_\lambda^{(b)}). \qquad [2.20]$$

The detector response of the combination of sources is thus the sum of the detector responses of each source. The equation affords a means of dimensional reduction: full disclosure of the spectral power distribution of a set of sources is not required in modeling the collective response of a system of superpositioned sources. Computationally, the left-hand side of [2.20] may be evaluated by summing the precomputed scalar detectors appearing in the right-hand side. This is demonstrated later in §3.1.2.

This precomputation also suggests the substitution of alternate spectral power distributions having identical detector response(s) into any term appearing in the right-hand side. These represent lights which behave identically under the action of mixing in reference to the predetermined detector set. When the detector set is the tristimulus vectors the alternate sources are visually indistinguishable. Such sources are *metameric.* Equation [2.20] confirms the laws of additive colour mixture as set forth by in the last century (Grassman, 1853). It are also early statement of the law of linear algebra presented in the context of colorimetry (Krantz, 1975).

### 2.2.6 Ray Reflection (Subtractive Mixing)

A ray reflected from or transmitted through coloured media has its spectral power distribution modified by selective attenuation of energy at each wavelength $\lambda$ by the dimensionless factor $\rho(\lambda)$. The reflectance equation

$$\Phi_\lambda^{(r)} = \rho(\lambda)\ \Phi_\lambda^{(i)} \qquad \text{[ray reflection]} \qquad [2.21]$$

is identical to [2.2] except $\rho(\lambda)$ is multiplied through. This models the *Helmholtz reciprocity* (Clarke and Parry, 1985), indicating that reflectance is an intrinsic surface property not functionally dependent upon the distribution of energy of the incident ray. This is the invariance suggested in the thesis's preface. Composition of (a set of) reflectances is the natural phenomena by which colour is most often observed in everyday settings. For instance, a solar ray is successively reflected by objects in an outdoor setting until reaching the retina. Algebraically,

$$\Phi^{(\mathrm{r})}(\lambda) = (\rho^{(\mathrm{S_a})}(\lambda)\rho^{(\mathrm{S_b})}(\lambda)\cdots\rho^{(\mathrm{S_z})}(\lambda))\Phi^{(\mathrm{i})}_{\lambda}. \qquad [2.22]$$

Computation of such systems underpins graphics rendering packages which strive for photorealism (Cohen and Wallace, 1993). The term is recently coined (Hall, 1983), (Meyer and Greenberg, 1986) and describes graphics software which applies methods of ray-tracing (Whitted, 1980), (Glassner, 1989) and in particular, scene radiosity or scene radiosity (Cohen et al., 1986) to describe overall illumination and effects such as partial shadows. In the case of ray tracing, a "reflection" procedure[7] is invoked each time a selected ray impinges upon some surface within the scene (a spatial model). Generally, the incident ray is transferred across a sequential system of surfaces $S_a \odot S_b \odot \cdots \odot S_z$ giving rise to a final reflected ray seen in [2.22] above.

The action of the system of reflectances may be modeled as the action of a single hypothetical reflectance appearing as the parenthesized product above. That is,

$$\rho^{(\mathrm{S_a}\odot\mathrm{S_b}\odot\cdots\odot\mathrm{S_z})}(\lambda) = \rho^{(\mathrm{S_a})}(\lambda)\rho^{(\mathrm{S_b})}(\lambda)\cdots\rho^{(\mathrm{S_z})}(\lambda), \qquad [2.23]$$

defines the composite system. The composition operator "$\odot$" forms the new reflectance; it is represented by multiplication of the continuous functions appearing on the right-hand side. The system reflectance is physical because it satisfies the

[7] See in particular the pseudo-Pascal appearing in §4.1.4.

reflection law $0 \leq \rho(\lambda) \leq 1$. That is, the product of values lying within the closed unit interval $[0 \ldots 1]$ remains within this interval. The presence of this physical constraint has extensive consequences on any mathematical models which attempt to approximate physical reflectances, as seen in the next chapter.

Note that the incident and reflected sources of illumination appearing in [2.22] may be dropped in [2.23], because the composite reflectance may be computed *without further reference* to any incident source $\Phi_\lambda^{(\mathrm{i})}$. This suggests a method of simultaneous radiometric and spatial rendering which makes no reference to any source of illumination while rendering "reflective" pixels now possessing both both spatial and radiometric coordinates. This extends the class of radiosity-based framebuffer methods of emerging importance described beginning at §8.2.

The multiplicative nature[8] of systems of reflectance account for the difficulty in producing practical computational models of subtractive colour. Linear factorization analogous to [2.20] does not hold because in general the detection of a multiply-reflected ray of light is not a sum-of-terms form. As an example, the $Y$ tristimulus coordinate of a system of surfaces is computed by substituting [2.23] into the detection equation [2.13] giving

$$Y = \int \Phi_\lambda \rho^{(\mathrm{a})}(\lambda)\rho^{(\mathrm{b})}(\lambda) \cdots \rho^{(\mathrm{z})}(\lambda)\bar{y}(\lambda)d\lambda. \qquad [2.24]$$

This equation suggests no simple factoring. Tradition methods offer two workarounds. The first replaces a measure of relative transmission with logarithmic density: $\delta(\lambda) = -\log_{10}[\rho(\lambda)]$. Since $\log(x_1 \times \ldots \times x_z) = \log(x_1) + \ldots + \log(x_z)$, a sum of terms form is created, as when modelling a system of transmissive objects such as a stack of optical filters or photographic negatives. Though not without

[8] Multiplicative colour would be a more appropriate term. Subtractive is applied in the same colloquial sense that "10% off" means "0.9×".

representational virtue (Evans, Hanson and Brewer, 1953), ray summing operations as defined in [2.19] must now involve antilogarithms. The simultaneous additive and multiplicative nature of ray compositions has not been reconciled. Instead the computational ease has been shifted toward the multiplicative.

Alternately, the integrand appearing in [2.23] may be simplified should either the responsivity weighting factor or reflectance function terms take a simple form. Specifically, functions which are zero (one) everywhere outside a subrange of definition can be either subsumed into the product or will drive the entire equation to zero outside their limits of integration. That is, the interactions of the terms in the integrand are decoupled. Pursuit along these lines leads to exploration of "model" reflectance functions such as ideal cyan, magenta and yellow, which define idealized high-pass, notch and low-pass filters and whose reflective combinations may be modeled using a simple algebra (Foley et al., 1990). Generalization of this technique considers sets of "tine" filters which are bimodel in value and which form orthogonal sets (§8.3.1). Such idealized surfaces are hard to achieve in practice, although physical models approximating their behaviour are valuable. Further removed are predetermined model reflectances which are both continuous and orthogonal. These need no longer satisfy [2.2] and therefore no longer have associated physical counterparts. The definition and use of such reflectance bases is a central subject of Chapter 4.

## 2.3 Colour Spaces

The generation equation [2.9] suggests the modeling of a ray's spectral power distribution using a linear space of representation. Conceptually, the scalar $\alpha$ describes the linear contribution of a continuous distribution $\phi_\lambda$, which need not describe

a physical source[9]. Combined with the equation of superposition of light [2.19], a set of values $\alpha_1 \ldots \alpha_n$ describe the linear combination of a set of sources. The accompanying $n$ independent elements $\phi^{(1)} \cdots \phi^{(n)}$ then define a basis spanning an $n$-dimensional space. The physical colour (distribution) $\Phi_\lambda^{(i)}$ associated with any element $\phi^{(i)}$ may be determined by setting $\alpha_i = 1$.

The CRT model may be treated in terms of a three-dimensional colour space of illumination. The three colour components of a pixel $\mathbf{p}$ govern the excitation of three phosphors (having normalized distributions $\phi_\lambda^{(R)}$, $\phi_\lambda^{(G)}$, $\phi_\lambda^{(B)}$) forming a ray having continuous spectral power distribution $\Phi_\lambda^{(RGB)}$ (see also [3.1] through [3.3]). In loose terms, an *inflation* has taken place as the dimension of the parameters in $\mathbf{p}$ (three) give rise to a continuous distribution, albeit one having only three controlling degrees of freedom. In fact, the locus of points for all permissible $\mathbf{p}$ form a three-dimension (hyper) plane lying in a higher dimensional linear space of representation, as in the $n$-space whose $i$th coordinate records the value of $\Phi_\lambda^{(RGB)}$ at $\lambda = i$ for some discrete set of $n$ test wavelengths.

Conversely, the detection equation [2.12] provides a means of *deflation* which maps continuous spectral power distributions onto scalar values. As before, a set of independent basis elements is employed, here the set of detector responsivities. The tristimulus functions provide a useful example. Given a continuous spectral power distribution these determine a tristimulus vector $\mathbf{t}$ whose three coordinates $X$, $Y$ and $Z$ define a point in the tristimulus space: a three-dimensional colour space of detection whose points uniquely determine the realm of all perceivable colour.

[9]Allowing $\phi_\lambda$ to become negative admits the possibility of "black" light useful in the study of metamers in §3.1.7.

### 2.3.1 Representation and Dimensional Change

Colour spaces serve as an abstraction which facilitates the mapping between colour representations of varying dimension. When the representational models are linear the spaces are Euclidean, allowing the dimensional change to be phrased simultaneously in the language of linear algebra and (Euclidean) solid geometry. The adoption of a high-dimensional model (having vanishingly small approximation error when raised to the power of the continuum) does not imply increased computational complexity. To the contrary, the presence of such models supports proper use of simplified forms which map between low-dimensional representations without computational (but only analytic) recourse to the model's full generality on the continuum. This is made clear by the on-going CRT example as summarized in §3.1.4, wherein the mapping between pixels and tristimulus vectors is derived in terms of continuous functions while employing small, discrete linear systems.

The reflection equation [2.21] does not suggest the use of linear forms. Nonetheless, linear models may be applied to the representations of both spectral power distributions and of reflectance functions with success. A linear model which treats either form correctly must consider the disparity of dimension and representation in the many commonly encountered situations of colour, which are most often reflective. A thumbnail taxonomy appears in Figure 2.1.

### 2.3.2 The Colour Hierarchy

A hierarchy may be created to categorize the many additive and subtractive representations in common use. The colour hierarchy (Figure 2.1) is a directed graph whose nodes define colour models, annotated in bold-face. Nodes higher (lower) in the graph represent higher (lower) dimensional models. The top-most node

represents a physical surface whose colour is described by a continuous function; its direct descendent represents a continuous spectral power distribution (the surface under illumination). Progressing downward, the second descendent in vertical column describes the highest ($n$-) dimensional discretizations of the distributions which characterize this reflecting surface, as by spectroradiometry. This node forms the practical top of the hierarchy for any computational method. The leaves in the graph describe common models of colour having reduced, discrete dimension. The nodes appear as an S, R, or V, indicating that the associated "colour" is a property of a physical surface, a physical ray or a set of abstract values of detection or generation.

Edges define colorimetric operations which map between various representations. The name of the mapping operation appears in italic face. Downward edges define operations which discretize and reduce the dimension of colour representations and hence remove information (deflation). Upward edges define the generation of light, the production of physical surfaces or the computation of values of a higher-dimensional model given some discrete specification (inflation). These edges indicate an increase in the model's dimension of representation and not the creation of new information.

Lateral traverses indicate conversion between representations of similar dimension. The traverse may be lossless, as when mapping directly between RGB and XYZ trichromatic additive colour models (see also Appendix G), or lossy, indicated by a dotted line. Lossy paths generally represent desirable mappings for which no direct path exists. Instead, an indirect path may be established. To maximize the preservation of information, an upward traverse should precede a downward one. Of particular note is the mapping between RGB scanner information and CMY(K) print densities. Though of similar dimension, the presence of this path represents

Figure 2.1: The Colour Hierarchy

the open challenge in producing acceptable colour hardcopy and is the subject of extensive practical and theoretical investigation.

### 2.3.3 Linear Colour Space

The concepts and linear algebra underlying colour spaces of both rays and surfaces are now formalized. Notations are defined as used throughout the thesis. In the two chapters that follow, the use of linear systems are reviewed in the context of additive colour (Chapter 3) and reflective colour (Chapter 4).

A linear colour space is spanned by a basis of independent functions, which may be linearly combined to represent spectral power distributions or reflectance functions. That is,

$$\Phi_\lambda = x_1\phi_\lambda^{(1)} + x_2\phi_\lambda^{(2)} + \ldots + x_m\phi_\lambda^{(\mathrm{m})} \qquad \text{[additive model]} \qquad [2.25]$$

for additive colour, and

$$\rho(\lambda) = x_1\beta_1(\lambda) + x_2\beta_2(\lambda) + \ldots + x_m\beta_m(\lambda) \qquad \text{[reflective model]} \qquad [2.26]$$

for reflective colour.

Whereas [2.25] describes a physical system of sources, [2.26] does not describe a physical system. This represents a point of departure. In additive systems,

> "The outstanding property of a chromaticity diagram is that the representative point of an additive mixture of two colour stimuli lies on the straight line passing through the chromaticity points corresponding to the constituents of the mixture." (Wyszecki and Styles, 1982c)

In reflective colour space an analogous concept involving physical reflectances (i.e., points) in linear combination does not exist. Instead, the physical reflectances corresponding to points in the space must be explored by other means. This is treated at length in §4.1.

When the basis elements are independent, the colour coordinates $(x_1, \ldots x_m)$ describe a unique point in $m$-dimensional linear space, called a *colour space.* The space is *Euclidean* when an inner product is introduced (Apostol, 1969a), allowing measurement of both distances and angles. The inner product is used to establish orthogonality among the basis elements. Orthogonal basis sets simplify means of computation and methods of least-squares approximation within colour spaces. Moreover, Euclidean colour spaces of low dimension suggest the construction of physical colour models used to visualize the spaces. Three-dimensional additive physical models based on the trichromatic nature of human vision are common. The colour reflectance models explored in the following chapter extend this to subtractive colour at three and higher dimensions.

A basis function need not describe a physically realizable colour satisfying either the illumination law [2.1] or reflection law [2.3]. However, a suitable linear combination of basis elements will. That is, not all points in colour space need define physical colour. Conversely, a continuous spectral function $\Phi_\lambda$ or $\rho(\lambda)$ need not lie precisely at a point in a colour space. This is treated later.

### 2.3.4 Discretization and Dimensional Considerations

Any continuous (spectral) distribution may be represented as a discrete $n$ element column vector, with $n$ the sampling dimension, as when created by spectroradiometry. For instance, reflectance functions $\rho(\lambda)$ become *reflectance vectors:* $\mathbf{r}$. Re-

flectance spectra are band limited (Bracewell, 1978); researchers (Stiles, Wyszecki and Ohta, 1977), (Wyszecki and Styles, 1982a) have examined the frequency dependence on the amplitude of the coefficients of their associated Fourier spectra. Simply put, increasing the number of samples $n$ (across the fixed domain of visible wavelength) increases the sampling frequency and hence reduces the approximation error between discrete and continuous representations. Choosing an arbitrarily large $n$ drives the error to zero as no aliasing effects are present as an information theoretic consequence of band limitation (Panter, 1965). The continuous distribution $\rho(\lambda)$ may be reconstructed as required by methods of piecewise-linear interpolation[10].

Using this linear model, any ray or surface having distribution $\Phi_\lambda$ or $\rho(\lambda)$ can be characterized exactly using a discrete $n$-dimensional linear space, with $n$ chosen arbitrarily large. That is, $n$ is the dimension of discretization of the illumination space. In empirical practice, $n$ takes on the sampling resolution of the instrumentation used in producing all data sets under study. Representation within this "parent" linear space is considered exact. That is, any distribution defines a unique point in this $n$-dimensional colour space, which is spanned by an independent basis. The $n$ components of this unique point (vector) $\mathbf{x}$ are the *colour coordinates* of the distribution. The data presented throughout this thesis was sampled at a spacing of $1nm$ ($m = 381$ across the domain $370nm \leq \lambda \leq 750nm$), though culling to $5nm$ ($m = 76$) was occasionally employed for computational speedup. These sampling resolutions are well within accepted colorimetric practice.

A coordinate vector of reduced dimension $m < n$ may also represent a distribution, hence a colour. For instance, the *reflectance coordinates* $\mathbf{x}$ in $m$-space define a discrete reflectance $\mathbf{r}$ in $n$-space. As $m \to n$, the reflectance coordinates define the

[10]This is one of two interpolation methods prescribed by the CIE for use with reflectance data (Commission Internationale D'Clairage, 1986).

reflectance vector exactly (to within a change of basis). Without loss of generality, the basis may be set to the identity. In this case, $\mathbf{r} = \mathbf{x}$ under the "tine" element basis $\beta_i(\lambda)$ in which each $i$th element is nonzero at one narrow band sampling wavelength. Under the tine model, a conventional spectroradiometer (which produces a discrete representation of $\rho(\lambda)$ by measuring flux at successive wavelengths) is better regarded as an instrument which produces a set of reflectance coordinates in reference to the orthogonal tine basis. Both conceptualizations are illustrated in Figure 2.2.

This decoupling of reflectance coordinates and reflectance vectors allows $m$ to vary independent of $n$. The conceptual model also suggests a new style of spectroradiometery. The conventional instrument synthesizes the tine detector basis using an optical slit and diffraction grating which together produce linear dispersion in spectra as a function of wavelength. The new model admits the possibility of a device which records reflectance coordinates under a more general detector basis. Subsequent application of [2.16] allows the output to be presented in the conventional form[11].

### 2.3.5 Notations and Matrix Algebra

Continuous functions describing distributions appear in the traditional notation[12] as Greek letters. Discrete counterparts are column vectors, appearing as bold lowercase type. The corresponding Roman letter is employed where possible. Distributions are in turn expressed as linear combinations of either continuous or discrete

[11] As an example, instrumentation using a detector set based upon the Walsh/Hadamard functions and enjoying enhanced low-light sensitivity are now commercially available.

[12] Following a common usage (Wyszecki and Stiles, 1982) which upholds prior convention (Nicodemus et al., 1977), (Nicodemus, 1978), (Spencer and Gaston, 1975).

Figure 2.2: Reflectance Vector Models

| | Distributions | | Bases | | Colour |
|---|---|---|---|---|---|
| | cont. | disc. | cont. | disc. | Coordinates |
| Additive | $\Phi_\lambda$ | $\mathbf{s}$ | $\phi^{()}_\lambda$ | $\mathbf{s}^{()}$ | $\mathbf{p}$ |
| Reflective | $\rho(\lambda)$ | $\mathbf{r}$ | $\beta(\lambda)$ | $\mathbf{b}$ | $\mathbf{x}$ |

Table 2.1: Variable Notation (Linear Algebra)

basis elements, suggested by $\mathbf{b}$ and $\beta$. The symbols used are tabulated in Table 2.1.

Most generally, colour coordinates are presented as a $\mathbf{x}$, though the letter $\mathbf{p}$ is chosen to indicate *pixel*, specific to additive systems. (In this case the components of pixel $\mathbf{p}$ also take on discrete values and are most often hardware integers.) Finally, discrete (column) vectors may be augmented to form arrays, in which case an upper-case equivalent is substituted. For instance, $\mathbf{B}$ is the $m \times n$ basis formed by augmenting $\mathbf{b}_1 \ldots \mathbf{b}_m$; $\mathbf{D}$ and $\mathbf{G}$ are respective detection and generation matrices.

Matrix multiplication allows the simultaneous evaluation of inner (dot) products to take place among column elements. In this setting one matrix is transposed, so that (for instance) the product of $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ contains elements $(i, j)$ which are the dot (inner) product of the $i$th column of $\mathbf{A}$ with the $j$th column of $\mathbf{B}$, across the range of conforming (row) wavelengths. This form arises most often when the columns of one matrix represent discrete detector responsivities, the other a spectral power distribution. For instance, the parallel action of a spectral power distribution $\mathbf{s}$ on three detectors whose discrete responsivities $\mathbf{d}_1 \ldots \mathbf{d}_3$ form the three columns of matrix $\mathbf{D}$ may be represented as

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_1 \\ v_3 \end{bmatrix} = \mathbf{D}^{\mathrm{T}}\mathbf{s} \qquad [2.27]$$

in which $\mathbf{v}$ is the matrix of output values. Note that both $\mathbf{D}$ and $\mathbf{s}$ necessarily conform in row number because the discrete sampling wavelengths must match by row number. Similarly, matrices which conform in both row and column number may be summed. For instance, when $\mathbf{D}$ contains the tristimulus vectors, the linear operations on which Grassman's laws of colour are based can be expressed as:

$$\mathbf{D}^{\mathrm{T}}(\mathbf{s} + \mathbf{s}') = \mathbf{D}^{\mathrm{T}}\mathbf{s} + \mathbf{D}^{\mathrm{T}}\mathbf{s}'. \qquad [2.28]$$

In this guise $\mathbf{D}$ serves as the linear operator[13], $\mathbf{s}$ and $\mathbf{s}'$ its operands. Note in particular that the operator notation used to define detector summation [2.16] is easily expressed as

$$(\mathbf{D}_a + \mathbf{D}_b)^{\mathrm{T}}\mathbf{s} = \mathbf{D}_a^{\mathrm{T}}\mathbf{s} + D_b^{\mathrm{T}}\mathbf{s}, \qquad [2.29]$$

because either the matrix or vector may be distributed across the sum.

### 2.3.6 Procedural Methods

The use of column vectors further facilitates the off-line recording of large spectroradiometric data sets. Since the matrices described typically employ a small ($< 8$) number of columns but a large number of rows, matrices may be conveniently represented as tables in ASCII format and manipulated using text editors and other conventional tools. Because versions of MATLAB and related software could not accommodate the largest data sets produced[14], a custom C-language library supporting general matrix and colorimetric operations was created (§E.1). The library supports runtime determination of matrix size, provides conformality checks, and

[13] That is, which can be distributed or associated in linear expressions. A common example is differentiation: $d(x + y) = d(x) + d(y)$ where $d$ is the differential operator and $x$, $y$ are functions.

[14] The full singular value decomposition upon a $4069 \times 512$ data set was the largest calculation attempted. Note that one row of the matrix exceeds the line input length of most software.

allows an optional wavelength annotation to appear in any data set as a prefacing zeroeth column. This last feature supports a mechanism for linear (row) interpolation of matrix data whose prefacing wavelength value is strictly monotonic, thereby extending the concept of conformality and making possible a large software suite of colorimetric tools of general use (See also §E.2). The radiometric procedure was similar to methods previously employed by the author in conducting related work (Paeth, 1989a), (Paeth, 1989b). (Empirical tables employed are reproduced in Appendix G.)

## 2.4 Summary and Conclusions

An overview of the basic physics of colour has been presented. A simple taxonomy of the traditional methods of colour representation suggests a dimensional hierarchy. Methods of linear algebra have been developed. In particular, the *colour space* and related concepts have been introduced to provide a consistent framework of colour description which operates across an arbitrary range of dimension. At the highest dimension $n$, the framework subsumes the traditional sampling model, at reduced dimension $m$ a hierarchy of lower-dimensional models becomes possible. The methods and notations presented easily describe the mixing of light and provide the basis of exploration of additive colour spaces and the nearly analogous *reflectance space* presented in the following two chapters.

# Chapter 3

# Additive Colour Space

The concepts of additive colour space may now be formalized based upon the methods which conclude the previous chapter. The next section reviews traditional (additive) colorimetric methods such as metamerism in terms of the thesis's formalisms. The second section describes additive colour space in geometric terms. Though much of this portion of the formalization is review[1], both the matching of colour and geometric solids have close analogs of primary importance to the reflectance spaces described in the next chapter.

## 3.1 Linear Spaces of Additive Colour

This section considers the mapping of CRT pixels $\mathbf{p}$ onto tristimulus values $\mathbf{t}$. The mapping is in three steps. First, the illumination coordinates of the pixel $\mathbf{p}$ are *inflated* from 3-space into the $n$-space spanned by a basis of monochromatic sources.

[1]Codification of additive colour in terms of linear algebra remains an active topic of contemporary research (Shafer, 1982), (Cohen, 1988).

This illumination space is conterminous with $n$-dimensional detection space of orthogonal tines, as provided by a conventional spectroradiometer. Deflation onto tristimulus space establishes the coordinates (components) of the vector $t$, completing the operation. The following sections summarize and derive the transformations.

### 3.1.1 Overview

A ray's generation, additive superimposition and ultimate detection may be modeled using linear algebra. A finite collection of $j$ independent sources can be controlled in linear fashion after the generation equation [2.9]. A $j$-dimensional vector records the contribution of each source, whose output is a continuous distribution. Likewise, the responses from a set of $k$ independent detectors may be recorded as a vector. Since this is also a linear operation, a linear transformation maps the $j$-vector onto the $k$-vector. The essential "colour" of the ray is recorded ultimately in terms of the detector values. As will be seen, all calculations may take place within the final $k$-dimensional space of representation without explicit recourse to the parent $n$-space. Both the $j$-space of independent sources and the $n$-space of all monochromatic sources are *illumination spaces.*

A well-defined set of (abstract) detectors model the space of perceptual colours. These are the tristimulus vectors which span the tristimulus space. They allow the mapping of an arbitrary, continuous spectral power distribution $\Phi_\lambda$ onto a *detection space,* here having three dimension. This establishes a correspondence between $\Phi_\lambda$ and a unique point $(X, Y, Z)$ lying in tristimulus space[2]. Since $\Phi_\lambda$ need not lie in the space spanned by the CRT phosphors, conditions of multiplicity and uniqueness

[2]The components named $X$, $Y$, $Z$ are both detected tristimulus values and spatial coordinates.

are also examined.

### 3.1.2 The CRT: An Additive Example

In the CRT example described below, rays having selected tristimulus vectors are produced through the superposition of light generated by three phosphors on a hypothetical, single-pixel frame buffer. The set of all combinations of phosphor output, and hence all displayable pixels on the CRT, comprise the *device gamut.* The gamut is often described as an abstract physical object: the locus of points in the additive space associated with presentable hues on the device[3].

Here, the pixel $\mathbf{p}$ is a 3-dimensional vector whose components $\mathbf{p}_{red}$, $\mathbf{p}_{grn}$ and $\mathbf{p}_{blu}$ are linear controls of light production, but having discrete value on a reduced set. (This is a central feature of previous work (Paeth, 1987a) and is re-addressed in the reflective domain in Chapter 8 and in Table 8.2. Abstract frame buffer models have been advanced (Higgins and Booth, 1986) and treated formally elsewhere (Fiume, 1989), (Paeth, 1990b).) The detector set (responsivities) are the tristimulus functions.

The distribution of the composite ray $\Phi_\lambda^{(\mathrm{CRT})}$ under control of a single pixel $\mathbf{p}$ is found by superposition of the rays produced by red, green and blue phosphors. These have normalized spectral power distributions $\phi_\lambda^{(\mathrm{red})}$, $\phi_\lambda^{(\mathrm{grn})}$ and $\phi_\lambda^{(\mathrm{blu})}$. The composite ray's distribution is then created using the generation model

$$\Phi_\lambda^{(\mathrm{CRT})} = \mathbf{p}_{red}\phi_\lambda^{(\mathrm{red})} + \mathbf{p}_{grn}\phi_\lambda^{(\mathrm{grn})} + \mathbf{p}_{blu}\phi_\lambda^{(\mathrm{blu})}, \qquad [3.1]$$

summed in linear combination. This equation becomes the integrand common to

[3]With well-chosen phosphors, the range of hue is typically much larger than is encountered in normal (reflective) scenes.

three detection equations [2.13], evaluated in parallel:

$$\begin{aligned} X &= \int (\mathbf{p}_{red}\phi_\lambda^{(\mathrm{red})} + \mathbf{p}_{grn}\phi_\lambda^{(\mathrm{grn})} + \mathbf{p}_{blu}\phi_\lambda^{(\mathrm{blu})})\ \bar{x}(\lambda)d\lambda \\ Y &= \int (\mathbf{p}_{red}\phi_\lambda^{(\mathrm{red})} + \mathbf{p}_{grn}\phi_\lambda^{(\mathrm{grn})} + \mathbf{p}_{blu}\phi_\lambda^{(\mathrm{blu})})\ \bar{y}(\lambda)d\lambda \\ Z &= \int (\mathbf{p}_{red}\phi_\lambda^{(\mathrm{red})} + \mathbf{p}_{grn}\phi_\lambda^{(\mathrm{grn})} + \mathbf{p}_{blu}\phi_\lambda^{(\mathrm{blu})})\ \bar{z}(\lambda)d\lambda \end{aligned} \qquad [3.2]$$

The above integrals contain a sum-of-terms form which may be regrouped:

$$\begin{aligned} X &= \mathbf{p}_{red}\int \phi_\lambda^{(\mathrm{red})}\bar{x}(\lambda)d\lambda + \mathbf{p}_{grn}\int \phi_\lambda^{(\mathrm{grn})}\bar{x}(\lambda)d\lambda + \mathbf{p}_{blu}\int \phi_\lambda^{(\mathrm{blu})}\bar{x}(\lambda)d\lambda \\ Y &= \mathbf{p}_{red}\int \phi_\lambda^{(\mathrm{red})}\bar{y}(\lambda)d\lambda + \mathbf{p}_{grn}\int \phi_\lambda^{(\mathrm{grn})}\bar{y}(\lambda)d\lambda + \mathbf{p}_{blu}\int \phi_\lambda^{(\mathrm{blu})}\bar{y}(\lambda)d\lambda \\ Z &= \mathbf{p}_{red}\int \phi_\lambda^{(\mathrm{red})}\bar{z}(\lambda)d\lambda + \mathbf{p}_{grn}\int \phi_\lambda^{(\mathrm{grn})}\bar{z}(\lambda)d\lambda + \mathbf{p}_{blu}\int \phi_\lambda^{(\mathrm{blu})}\bar{z}(\lambda)d\lambda \end{aligned} \qquad [3.3]$$

which is a traditional colorimetric form.

The detected response (an inner product) of each source against each detector responsivity may now be precalculated. For instance, $\mathrm{x_{red}} = \mathrm{D}^{(\bar{\mathrm{y}})}(\phi^{(\mathrm{red})}) = \int \phi_\lambda^{(\mathrm{red})}\bar{\mathbf{y}}(\lambda)\mathrm{d}\lambda$ defines the amount of red registered by the detector with discrete (tristimulus) responsivity $\bar{\mathbf{y}}$ and discrete distribution $\phi_\lambda^{(\mathrm{red})}$. The nine terms are gathered in a $3 \times 3$ matrix, giving

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \mathrm{x_{red}} & \mathrm{x_{grn}} & \mathrm{x_{blu}} \\ \mathrm{y_{red}} & \mathrm{y_{grn}} & \mathrm{y_{blu}} \\ \mathrm{z_{red}} & \mathrm{z_{grn}} & \mathrm{z_{blu}} \end{bmatrix} \begin{bmatrix} \mathbf{p}_{red} \\ \mathbf{p}_{grn} \\ \mathbf{p}_{blu} \end{bmatrix} \qquad [3.4]$$

The advantage of this formulation is that additive colour models need carry neither the spectral power distributions or detector responsivities in continuous form. The equation may be written compactly as

$$\begin{aligned} \mathbf{t} &= \mathbf{C}\mathbf{p}, \\ &= [XYZ]^{\mathrm{T}}. \end{aligned} \qquad [3.5]$$

in which the tristimulus vector $\mathbf{t}$ is derived from pixel vector $\mathbf{p}$ using a matrix $\mathbf{C}$ whose elements are precalculated from the spectral power distributions and detector responsivities appearing in [3.2]. In practice, $\phi_\lambda^{(\mathrm{red})}$ is discretized by radiometry and becomes $\mathbf{s}_{red}$; $\bar{\mathbf{y}}$ is tabulated (§G.2) and the vector dot (inner) product substitutes

for the integrals. The vector notation clearly illustrates that the perceptual response of the sum of (pixel) sources $\mathbf{C}(\mathbf{p}+\mathbf{p}')$ equals the sum of the perceptual responses of each source $\mathbf{Cp}+\mathbf{Cp}'$, again obeying Grassman's second law concerning the mixing of light.

Equation [3.5] maps pixel vector $\mathbf{p}$ having *illumination coordinates* onto the tristimulus vector $\mathbf{t}$ under the coefficient matrix $\mathbf{C}$. Geometrically, $\mathbf{C}$ describes a linear transformation of $\mathbf{p}$ onto the tristimulus space. Conversely, $\mathbf{C}^{-1}$ may be regarded as the mapping of the abstract detectors $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$ onto the space of $\mathbf{p}$ spanned by the distributions $\phi_\lambda^{(\mathrm{red})}$, $\phi_\lambda^{(\mathrm{grn})}$ and $\phi_\lambda^{(\mathrm{blu})}$.

### 3.1.3 $N$-dimensional Illumination Space

By use of sufficiently large n, an $n$-dimensional illumination space records exactly (to the limits of instrumentation) any continuous spectral function. For instance, the pixel $\mathbf{p}$ (a 3-vector) defines the point in the parent $n$-dimensional illumination space (spanned by the orthogonal "tine" basis of monochromatic sources). This is an instance of generation, written as

$$\mathbf{s} = \mathbf{Gp}, \qquad [3.6]$$

in which $\mathbf{s}$ (an $n$-vector) is the composite spectral power distribution and $\mathbf{G}$ the $n \times 3$ generation matrix formed by augmenting the column vectors $\mathbf{s}^{(\mathrm{red})}$, $\mathbf{s}^{(\mathrm{grn})}$, and $\mathbf{s}^{(\mathrm{blu})}$ which discretize the spectral power distributions of each phosphor.

This $n$-space of sources is isomorphic with the $n$-space of detectors spanned by monochromatic detectors. That is, the coordinates of a point in this space may be simultaneously regarded as the generation values against the tine basis or monochromatic sources, or as the detector responses against this orthogonal set of

detector responsivities. (Here $\Phi_\lambda$ and $\mathbf{s}$ are similar, as are $\phi_\lambda^{()}$ and $\mathbf{s}^{()}$, as per Table 2.1.)

Simultaneous detection may be phrased in a similar notation:

$$\mathbf{t} = \mathbf{D}^{\mathrm{T}}\mathbf{s}. \qquad [3.7]$$

Here $\mathbf{D}$ contains columns which are the detector responsivities. When these colours are the tristimulus vectors[4], a tristimulus value $\mathbf{t}$ is then derived from the discrete spectral power distributions $\mathbf{s}$. The mapping of a frame buffer pixel (vector) $\mathbf{p}$ onto the trichromaticity vector $\mathbf{t}$ is therefore

$$\mathbf{t} = \mathbf{D}^{\mathrm{T}}(\mathbf{G}\mathbf{p}). \qquad [3.8]$$

so that

$$\mathbf{C} = \mathbf{D}^{\mathrm{T}}\mathbf{G} \qquad [3.9]$$

provides the means for computing the constants which form the elements of the matrix[5] $\mathbf{C}$ as in [3.5].

Since the tristimulus vectors do not form an orthonormal set, the mapping from pixel coordinates to tristimulus coordinates may not be represented as a rigid rotation of either coordinate system. Nonetheless, exploration of the geometry of additive colour spaces is instructive and is treated in §3.2. The geometric models are particularly useful in describing reflectance spaces, studied in Chapter 4.

[4] These functions are plotted in Figure 3.1.

[5] A representative set of $3 \times 3$ RGB conversion matrices and their inverses appear in Appendix F. They are derived both from manufacturers' specifications and from international standards.

### 3.1.4 Inflation and Deflation (Additive)

The parent $n$-space of representation is the common superset of all lower dimensional representations. As suggested by the Colour Hierarchy (Figure 2.1), it is through this common ancestor that all representations may be related. Equation [3.6] losslessly *inflates* the $j$-dimensional illumination coordinates of pixel $\mathbf{p}$ onto a larger $n$-vector of coordinates on the parent $n$-space of monochromatic sources. This space is conterminous with the $n$-space of monochromatic detector responsivities. The detection operation [3.7] *deflates* this $n$-dimensional vector onto a smaller $k$-dimensional vector whose components represent discrete detector values. As seen in [3.8], the $n$-dimensional dependencies are removed, leaving merely the $k \times j$ transformation matrix $\mathbf{C}$.

When $j = k$ the number of independent sources and detectors are in agreement. Then $\mathbf{C}$ is square and invertible, allowing for a lossless one-to-one mapping between XYZ tristimulus values and RGB pixel (illumination) coordinates. This appears in the Colour Hierarchy as the double arrow solid line between the RGB and XYZ leaves appearing at the bottom centre of Figure 2.1.

When the dimensions of the source and detector set are dissimilar the transformation matrix $\mathbf{C}$ is rectangular and the system is either over- or under-determined. Equations [3.6] and [3.7] may again be used for conversion by way of inflation onto the parent $n$-space followed by deflation onto the target space. In contrast to out of gamut cases, a desired set of detector values may prescribe multiple, linearly related solutions. This is the subject of the next section.

### 3.1.5 Metamerism of Light

Light having dissimilar spectral power distribution may evoke an identical perceptual response. These are called *metamers.* The notation

$$C_a \sim C_b \qquad [3.10]$$

indicates that the colours $C_a$ and $C_b$ are metameric matches.

This has practical value, for by this mechanism a CRT (for example) having no "yellow" phosphor can create yellow light by the additive mixing of rays from the red and green phosphors. These have tristimulus values which when combined in positive proportions produce the tristimulus value of a light whose spectral power distribution elicits a "yellow" sensation.

Metamerism occurs when distinct spectral power distributions have coinciding points within tristimulus space. In a more general sense, metamerism occurs within a detection space of dimension $k < n$ when the points of projection of two dissimilar distributions $\Phi^{(C_a)}$ and $\Phi^{(C_b)}$ coincide. Algebraically, [3.10] holds when

$$\begin{bmatrix} D_1(\Phi^{(C_a)}) \\ \vdots \\ D_k(\Phi^{(C_a)}) \end{bmatrix} = \begin{bmatrix} D_1(\Phi^{(C_b)}) \\ \vdots \\ D_k(\Phi^{(C_b)}) \end{bmatrix}, \qquad [3.11]$$

as vectors are identical when all components are equal. This demonstrates an identical behaviour of two lights – whose colours $C_a$ and $C_b$ have distinct distributions – in the context of a common set of detectors $D_1 \ldots D_k$.

The liability of this approach lies in the dissimilarity of distribution of two metameric sources. Their difference may not be accounted for by the detector set, yet can be inadvertently revealed in the presence of selective filtering. For instance, a narrow-band yellow filter transparent around 580nm will pass energy

from a monochromatic source of matching wavelength, yet the yellow formed on the CRT (showing a bimodal distribution with energy peaks near 550nm and 600nm and identical in appearance) will be attenuated. Unfortunately, reflective materials exhibit preferential reflectivity: the application of a linear space of reduced dimension, as in [3.4], no longer holds. This reveals a common misapplication of additive methods: the "colour" of the source is not sufficiently recorded by its tristimulus (or RGB) coordinates should interactions with surfaces also be taken into account.

### 3.1.6 Computation of Metamers

The computation of matching coordinates of two sources provides an instructive example in both metamerism and discrete matrix methods. The discrete spectral power distribution of a source of illumination $\mathbf{s}$ is given as a column vector. The pixel values $\mathbf{p}$ (whose coordinates $R$ $G$ and $B$ lie in a space of illumination) are sought, whose associated and distinct spectral power distribution has matching appearance. Parallel detection is represented by $\mathbf{D}$, which is the three tristimulus vectors in discrete form[6].

Premultiplication of $\mathbf{s}$ by $\mathbf{D}$, the $n \times 3$ tristimulus detection matrix, gives the tristimulus values $\mathbf{t}$. An RGB-to-XYZ conversion matrix $\mathbf{C}$, based on the manufacturer's specification (Appendix F) defines the mapping from $\mathbf{p}$ onto $\mathbf{t}$ as defined in [3.5]. Its inverse, $\mathbf{C}^{-1}$, converts from XYZ to RGB coordinates, allowing the computation of the frame buffer's RGB values as the column vector $\mathbf{p}$:

$$\mathbf{p} = \mathbf{C}^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{s}. \qquad [3.12]$$

[6]These are tabulated by the CIE to four-place accuracy at 5nm intervals and are reproduced in §G.2.

When the spectral power distributions of each phosphor are known (less common in practice), the elements of matrix $\mathbf{C}$ may be inferred by computing the tristimulus coordinates of each of the phosphors directly:

$$\mathbf{C} = \mathbf{D}^{\mathrm{T}}\mathbf{G}. \qquad [3.13]$$

Here $\mathbf{G}$'s columns are the discretized spectral power distributions of its phosphors. (The letter $\mathbf{G}$ is chosen as phosphors produce light by linear generation, after equation [2.9]) Substitution into the previous equations yields:

$$\mathbf{p} = (\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{s}. \qquad [3.14]$$

which describes the pixel RGB values needed to reproduce a matching, metameric source. Note that $(\mathbf{D}^{\mathrm{T}}\mathbf{G})$ is square and hence potentially invertible. Its size is $3 \times 3$ indicating the dimensional match between the number of distinct phosphors and the number of cones. That is, creation of metameric light is neither under- nor over-determined. Because neither $\mathbf{D}$ nor $\mathbf{G}$ are square, the inverses of the terms appearing within the subexpression do not exist. That is, the left hand factor may not be rewritten by applying the identity:

$$(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1} = (\mathbf{G}^{-1}\mathbf{D}^{\mathrm{T}})^{-1} \qquad [3.15]$$

Were this not so, substitution into the previous equation would allow the full cancellation of $\mathbf{D}^{\mathrm{T}}$, the presence of the detector matrix would be fully removed, and "metameric match" would become "agrees at all wavelengths".

Given an arbitrary source $\mathbf{s}$, the spectral power distribution of its metameric match $\tilde{\mathbf{s}}$ (here " ˜ " signifies "approximating") may be computed immediately by post-multiplication with $\mathbf{G}$, which generates the ray distribution from the tristiumulus values implicit in the terms $(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}$:

$$\tilde{\mathbf{s}} = [\mathbf{G}(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}]\ \mathbf{s}. \qquad [3.16]$$

The bracketed expression is both square and singular. It is discussed again in §3.1.8; analogous reflectance forms appear in §4.1.7.

### 3.1.7 Metameric Equality

Two sources $\mathbf{s}_1$ and $\mathbf{s}_2$ are metameric matches $\mathbf{s}_1 \sim \mathbf{s}_2$ when their detected values agree under a common detector set $\mathbf{D}$. The detector set $\mathbf{D}$ provides a colour space of reduced dimension $k$ in which the match occurs [3.11]. When $k \geq 3$ and the detector set spans the tristimulus space, perceptual matching becomes possible. Still, sources of illumination may be perceptually distinct in the presence of reflective surfaces. When $k = n$ the matching is complete: the spectra of both $\mathbf{s}_1$ and $\mathbf{s}_2$ then agree at all discrete wavelengths.

Colour spaces provide models of reduced dimension which allow matching to occur. Given a suitably chosen set of $k$ detectors, the $n \times k$ matrix $\mathbf{D}$ defines the matching by

$$\mathbf{D}^{\mathrm{T}}\ \mathbf{s}_1 = \mathbf{D}^{\mathrm{T}}\ \mathbf{s}_2. \qquad [3.17]$$

As an example, $\mathbf{D}$ is formed by augmentation of the three tristimulus vectors. The (visually identical) metameric matching[7] between a given source $\mathbf{s}$ and its reproduction $\tilde{\mathbf{s}}$ on the CRT is then easily verified by comparing their respective chromaticity coordinates.

$$\mathbf{D}^{\mathrm{T}}\ \tilde{\mathbf{s}} = \mathbf{D}^{\mathrm{T}}\ \mathbf{s}. \qquad [3.18]$$

Substitution for $\tilde{\mathbf{s}}$ in the left-hand side and associating the first two matrices gives:

$$\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{s}} = (\mathbf{D}^{\mathrm{T}}\mathbf{G})(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{s} = \mathbf{D}^{\mathrm{T}}\mathbf{s}. \qquad [3.19]$$

[7] This is metamerism in its conventional sense.

The bracketed subexpressions are reciprocal and cancel by forming the 3×3 identity, proving the equality and hence the metameric match $\tilde{\mathbf{s}} \sim \mathbf{s}$. Thus, both $\tilde{\mathbf{s}}$ and $\mathbf{s}$ define a coincident point in tristimulus space defined by $\mathbf{D}$.

This may be restated as

$$\mathbf{D}^{\mathrm{T}}\ \tilde{\mathbf{s}} - \mathbf{D}^{\mathrm{T}}\ \mathbf{s} = \mathbf{D}^{\mathrm{T}}\ (\tilde{\mathbf{s}} - \mathbf{s}) = 0, \qquad [3.20]$$

indicating that the hypothetical adjusting source of light $(\tilde{\mathbf{s}} - \mathbf{s})$ is "black" in reference to the detector set. The sources of hypothetical spectral power distributions having tristimulus values $X = Y = Z = 0$ have been termed "null colour stimuli" by Wyszecki (Wyszecki and Styles, 1982c). These distributions are non-physical and may be superimposed onto the distributions of known sources to create metameric lights. More generally, such "black" lights are the set of $n - k$ independent basis elements which span the null space outside of the space of $k$ detectors. The reflectance analogs are the metameric surfaces of central concern in §6.1.

### 3.1.8 Dimensionality and Degrees of Freedom

The metameric source $\tilde{\mathbf{s}}$ derived from $\mathbf{s}$ has matching dimension (the column vector conforms) but is reduced in its degrees of freedom to $k$. The mapping equation

$$\tilde{\mathbf{s}} = [\mathbf{G}(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}]\ \mathbf{s}. \qquad [3.21]$$

may be generally expressed as

$$\tilde{\mathbf{s}} = \mathbf{M}\ \mathbf{s} \quad \text{with} \quad \mathbf{M} = \mathbf{G}(\mathbf{D}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{D}^{\mathrm{T}}. \qquad [3.22]$$

with $\mathbf{M}$ an $n \times n$ matrix. Again, neither the $\mathbf{G}$ nor $\mathbf{D}$ matrices may be combined with their inverses to simplify by cancellation. When $\mathbf{G}$ and $\mathbf{D}$ are identical (though

not necessarily orthogonal), then $\mathbf{M} = \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}$ is the *projection matrix* of $\mathbf{D}$ (Leon, 1990a).

In general, the matrix $\mathbf{M}$ is square and maps the source $\mathbf{s}$ having full degrees of freedom onto a source $\tilde{\mathbf{s}}$ of $n$ conforming rows but having reduced rank $k$ (three, in the case of the CRT example). It is clearly singular, because the existence of an inverse $\mathbf{M}^{-1}$ would provide a lossless means of reconstructing $\mathbf{s}$ from $\tilde{\mathbf{s}}$, an impossibility.

As expected, the product

$$\mathbf{s} = \mathbf{M}^2\mathbf{s} \qquad [3.23]$$

can be easily demonstrated. Thus $\mathbf{M}^i = \mathbf{M}$ for $i \geq 1$ indicating that once the point of projection is established within the colour space (and information lost to the null space) additional projections map onto the same point (no additional information is lost). Thus, the metameric match holds under transitive closure: the metameric match of a previous match is itself:

$$\tilde{\mathbf{s}} = \mathbf{M}^n\tilde{\mathbf{s}}. \qquad [3.24]$$

Singular, square matrices which resemble the identity in this fashion are called *idempotent.* Their use underpins modified techniques for constructions of orthogonal bases appearing in §5.1.4.

When the number of independent sources and detectors $k$ ($\mathbf{M}$'s rank) increases to the number of distinct wavelengths $n$, $\mathbf{M}$ becomes the identity and matching may then occur only when $\tilde{\mathbf{s}} = \mathbf{s}$ at all distinct wavelengths. Matching based on approximate agreement at all wavelengths is the basis of reflectance matching of colours under linear models of deflated dimension, the subject of Chapter 6.

## 3.2 The Geometry of Additive Colour Space

The $n$-dimensional domain of all discrete spectral power distributions and detector responses define additive colour spaces called *illumination spaces* or *detection spaces.* The inherent linearity of the mixing of light, plus the linearity of the cone detecting apparatus over a wide dynamic range (Wyszecki, 1981) make linear forms attractive. This is particularly true when each contributing source may be modeled by a linear system, whether by an inherently linear process, such as the gating of radiation using wire meshes[8] or by a linearized process, as on a CRT which has been linearized (gamma-corrected) using colour table lookup, as in [2.11].

Not all distributions correspond to an exact point in a colour space: $(\tilde{\mathbf{s}} - \mathbf{s}) = 0$ need not hold when $k < n$. More important, not all points in a colour space need define a physical source, since the latter are further subject to the physical law [2.1]. This section is concerned with identifying the locus of points in colour space which describe physical distributions, hence real colours. Most often the locus of points form a $k$-dimensional convex set, a *colour solid* or in this specific case, an *illumination solid.*

### 3.2.1 Illumination Solids

The $k$-dimensional illumination solid is composed of points whose corresponding distributions are everywhere positive. When $k \to n$, the solid is unique and represents all possible physical sources of illumination $\Phi_\lambda$. Because each basis element (tine) is non-negative, their relative contribution must be non-negative in any linear combination. Thus, the illumination coordinates of all points in the $n$-solid

[8] This is the classic means of attenuating a calibrated source without introducing spectral bias, modeling [2.9] with $1 - \alpha$ the coverage factor.

lie entirely within the first (hyper)octant of the space. Because the emissivity of any (monochromatic) source is unbounded, the corresponding coordinate may be arbitrarily large. This parent illumination solid in $n$-space is therefore cubelike, semi-open, infinite and convex. It may be defined as the convex hull of the ray bundle with each $i$th ray ($1 \leq i \leq n$) coincident with each $i$th coordinate axis and with all rays sharing a common terminal at the origin. Each $i$th ray defines the monochromatic tine source $\Phi_\lambda^{(\mathrm{i})}$ collinear with basis element $\phi_\lambda^{(\mathrm{i})}$. Convex closure establishes the set of interior points, whose associated spectral distributions are found by considering the hull vectors taken in positive linear combination. Since the axial ray bundle forming the hull is orthogonal, the $i$th component of the coordinates of any interior point defines the length along the related $i$th axis and hence the spectral energy present in the $i$th wavelength bin of the distribution.

Linearity indicates that any composite source $c$ having distribution $\Phi_\lambda^{(\mathrm{c})}$ and associated point $(X, Y, Z)$, describes a geometric ray (element) $\phi_\lambda^{(\mathrm{c})}$ whose locus of points $\alpha\phi_\lambda^{(\mathrm{c})}$ have illumination coordinates $(\alpha X, \alpha Y, \alpha Z)$. This model describes (linear) generation [2.19] in geometric terms. Setting $\alpha = 0$ for any source defines the black point, located at the origin. Thus, the locus of all physical sources in $n$-space is the ray bundle with common terminus at the origin, in which all rays lie in the first (hyper)octant.

The exterior surface of the volume may be found as the convex hull of the set of basis elements. For arbitrarily large $n$, the successive basis elements may be indexed in sequence using a spectral source "swept" across the domain of visible wavelength. Each element is then takes on unit energy within an infinitely narrow bandwidth (Dirac's delta function):

$$\phi_\lambda^{(\mathrm{test})} = \delta(\lambda_{test} - \lambda) \quad \text{with} \quad \int \delta(\lambda) d\lambda = 1 \text{ and } \delta(\lambda) = 0 \text{ when } \lambda \neq 0, \qquad [3.25]$$

thus producing the orthogonal elements $\phi_\lambda$ along the continuum which form the convex hull of this surface. Convex closure of this geometric object defines the (normalized) illumination solid, which may be expanded arbitrarily to define the unbounded solid formed as the hull of monochromatic sources $\Phi_\lambda$ under arbitrary scaling (increase of intensity). The convex closure allows the cyclic adjacency of the extreme sources of long (near infrared) and short (near ultraviolet) wavelength when the index passes from $n$ back to one, giving rise to a set of purple non-monochromatic sources which nonetheless lie along the solid's surface.

Lower dimensional illumination solids are subspaces of the parent $n$ solid and may be constructed through geometric projection (the inner product). In particular, the tristimulus space is formed when a linear change of basis establishes the tristimulus vectors as fictitious sources, with the remaining $n-3$ axes elided. Because projection of a convex solids preserves convexity (McKenna and Seidel, 1985), the locus of points in tristimulus space[9] is formed from the convex hull of the projected monochromatic sources. A representative drawing appears in (Wyszecki and Stiles, 1982), Figure 4(3.3.3). The tristimulus vectors are plotted in Figure 3.1.

The familiar chromaticity "horseshoe" of perceptual colour in tristimulus space may be regarded as an oblique projection of this $n$-hypercube. Because the tristimulus functions do not form an orthogonal set, the transformation is not a rigid rotation, it distorts the solid. However, the tristimulus functions are everywhere non-negative, preserving the solid's occupancy within the first octant. This offers geometric insight into the inequalities

$$X \geq 0, \qquad Y \geq 0, \qquad Z \geq 0 \qquad [3.26]$$

which hold for the tristimulus values of any physical source $\Phi_\lambda$. The inequalities

[9]These are termed *real stimuli,* see (Wyszecki and Stiles, 1982) p. 138.

Figure 3.1: The CIE Tristimulus Vectors (plots)

are easily verified algebraically by applying the detection integrals in [2.12] under the observation that spectral power distribution are everywhere non-negative [2.1].

Note that both (orthographic) projection and perspective projection are affine: they preserve linearity. Thus, geometric conditions of intersection, containment and set inclusion are preserved in any geometric object(s) created under these operations.

### 3.2.2 The Chromaticity Diagram

The chromaticity diagram is a well-known colorimetric tool used to encode the hue[10] of a source. The diagram encloses a convex and finite area. Analogous techniques related to solids of reflective colour are treated at length later.

The orientation of each ray in tristimulus space describes a normalized spectral power distribution and hence a perceptual hue, origin distance encodes brightness

[10]That is, perceived colour, without regard to overall brightness.

information. (This model ignores certain brightness induced hue change. These *additivity failures* include the *Purkinje shift* (Purkinje, 1823) caused by rod-dominant sensing at low levels, plus the *Bezold-Brüke effect* explained in part by saturation at high illumination levels, researched extensively (Purdy, 1931).) The brightness axis is easily removed by perspective projection onto the origin, thereby encoding merely the hue information in a two-dimensional *chromaticity diagram.*

The diagram's construction may be described geometrically. A plane of projection is erected which intersects each axis at a unit distance from the origin. The origin serves as a point of projection. This plane has defining normal vector $(1, 1, 1)$ and maps any point having homogeneous coordinates $(\alpha X, \alpha Y, \alpha Z)$ onto the point $(x, y, z)$. Algebraically, the mapping is

$$\begin{aligned} x &= X/(X+Y+Z) \\ y &= Y/(X+Y+Z)\,. \\ z &= Z/(X+Y+Z) \end{aligned} \qquad [3.27]$$

The system is overdetermined as the identity $x + y + z = 1$ holds save for the black source $\Phi_\lambda = 0$ (see below); the chromaticity of any non-zero equipotential source is $x = y = z = \frac{1}{3}$. Customarily, the $z$ is removed and the $x$ and $y$ chromaticity coordinates are plotted along the Cartesian axes of matching name. The diagram appears in Figure 3.2.

Note that the diagram is wholly contained within the triangle (0,0), (1,0), (0,1) which is the image of the plane of projection intersected with the first octant of tristimulus space (Figure 3.2(a)). The convex hull of the set of all possible sources $\Phi_\lambda$ forms the horseshoe, annotated by wavelength (in nanometers). This is the *spectral locus,* appearing as a continuous curve (solid) owing to the continuous nature of the tristimulus functions, though here discretized by plotting at the 5nm steps as originally tabulated by the CIE. (A revised table by the CIE removes noise

Figure 3.2: The CIE Chromaticity Diagram

while interpolating to a 1nm step size). Convex closure of the parent colour solid introduces the non-spectral *purple line* in the solid's projection. It appears along the bottom of the horseshoe (dotted) and joints points of extreme wavelength.

Points near the figure's boundary correspond to highly saturated colours; *white* sources lie in proximity to the projection of the equipotential source at $(\frac{1}{3}, \frac{1}{3})$. A set of three independent sources (e.g. the CRT phosphors) define a closed subset of distributions, called a (device) *gamut,* described in §3.2.3. Projection of the gamut's convex closure defines the triangular region which defines the domain of possible hue for the device, seen in Figure 3.2(b).

The *black point* $X = Y = Z = 0$ of projection is not defined on the chart as it lies at the point of projection. Algebraically, its computation by [3.27] results in a division by zero (see also the Preface). Geometrically the plane of projection passes through the origin, making the operation undefined.

In passing, it should be mentioned that the diagram's coordinate system is a simple bi-linear transformation (Birkhoff and MacLane, 1965). The ratio of two linear equations defines an affine transformation, preserving linearity. This explains both the triangular shape of the gamut and Wyszeki's observation regarding this preservation, quoted in §2.3.3. The perspective transformation common to computer graphics (Penna and Patterson, 1986) is a prime example of a bi-linear transformation and further illustrates that distance ratios are not preserved: the projection of a segment's midpoint is not the midpoint of the segment's projection (Newman and Sproull, 1971). This suggests the possibility of alternate planes of projection, based upon the arrangement of hue within the diagram. (For colour illustrations the reader is invited to refer to the plates immediately following page 616 in a widely available text (Foley et al., 1990).)

The *CIELUV* system defines the transformation

$$u' = \frac{4X}{X + 15Y + 3Z} = \frac{4x}{3 - 2x + 12y}, \quad v' = \frac{9Y}{X + 15Y + 3Z} = \frac{9y}{3 - 2x + 12y} \qquad [3.28]$$

used to approximate a *uniform colour space* in which a unit of distance in the space corresponds to a unit change in perceptual hue (Robertson, 1977). The geometric point and planes of projection are unimportant in this case. Instead, polynomials having small integer coefficients have been chosen, creating the *CIE 1976 UCS diagram* appearing in Figure 3.3. The library of colorimetric tools created in support of this research (Appendix E) supports both (x,y) and (u',v') style charts; the interactive geometric modeler whose examples appear throughout Chapter 7 support CIELUV charts as a display alternative to the chromaticity diagram.

An illustration[11] appears in Figure 3.3, based on the chromaticities of Ektachrome 5017 transparency film (ISO 64, in 35mm roll form) under source E for

[11]The same gamut appears in (Foley et al., 1990) as a chromaticity diagram under source A.

Figure 3.3: The CIE 1976 UCS diagram (CIELUV)

the primary and secondary hues. Note that cyan is seemingly out of gamut. This is a consequence of subtractive media: red and green are formed by the simultaneous activation of two dye couplers in the film's emulsion; cyan being a "primary" dye colour. This additivity failure is even more pronounced with reflective media.

### 3.2.3 Gamuts (Additive)

The set of all colours reproducible (in practice) in a colour space is called a *gamut.* For instance, the set of displayable pixels $\mathbf{p}$ describes the *device gamut* of the CRT under study within that illumination space. Since the CRT is a physical device, the gamut is contained within the tristimulus solid, and hence lies within (is a subset

The film chromaticity coordinates are tabulated in Appendix G. A sample spectrum is plotted in Figure 6.6.

of) the chromaticity diagram. In this sense, the space's colour solid represents the ultimate gamut achievable in theory by an ideal device having sufficient degrees of freedom.

For example, the CRT obeys a linear model of light production. Thus, its gamut in a three-dimensional space of illumination is the convex hull formed by three rays associated with the spectral power distribution of each phosphor[12]. The shape of the gamut is a triangular pyramid with apex at the origin. The origin represents the point of non-radiation (black). It opposes a base which represents distributions at the limits of physical light production under conditions of maximum drive voltage $V_0$. The limiting action of this base in no way alters the two-dimensional projection of the gamut onto the chromaticity diagram: it remains a triangle.

The hatched region lying outside the gamut of Figure 3.2(b) yet within the chromaticity diagram describes the points in tristimulus space having real stimuli not representable on the CRT. That is, a colour sensation might fall outside the CRT's range of presentable values. Such colours are *out of gamut* for the device. Their presence is suggested by the creation of forward and inverse mapping matrices $\mathbf{M}$ in [2.33]. Since the latter may contain negative elements (Appendix F), the inverse mapping of some (non-negative) tristimulus vectors onto pixel vectors may prescribe negative contributions from one or more sources, in violation of [2.1].

[12] This assumes that phosphor outputs may be blended continuously. In fact, the *palette* of CRT pixels is finite, as the RGB components are presented as discrete integers of chosen precision. Consequences and common misuses of this model are treated in (Paeth, 1990b).

### 3.2.4 Gamuts and Dimension

The volume of the space lying out of gamut may be reduced by increasing the number of elements of generation, such as by increasing the number of CRT phosphors. If an independent element is added having high spectral purity, the convex hull of the solid (and diagram) is increased. This is not done in practice as the gamut of colours achievable on a CRT are already much larger than the range of colours encountered in normal (reflective) settings. Were the gamut increased, the inflation of 3-dimensional tristimulus values onto a higher dimensional illumination space of pixel coordinates would be underdetermined. Instead, three sources are carefully chosen in practice which maximize the area of the gamut[13] appearing in Figure 3.1(b). These are the *primary colours of light* and are approximated by monochromatic sources of red (605nm) green (550nm) and blue (500nm) wavelength.

### 3.2.5 Other Additive Colour Spaces

Spaces of additive mixing are most often considered in three dimensions, so that (perceptual) metamers describe coincident points. A number of spaces having utility within the practice of computer graphics (Joblove and Greenberg, 1978), (Smith, 1978a) have become common. Because they are employed with CRTs, most assume maximum drive levels, often normalized to unity. As such, the spaces are closed; only their domain within the device's gamut is of interest. The spaces often called "triaxial" with a single character naming each axis by acronym. Most separate a

[13]Viewer discretion has advised brightness over hue purity: the Erbium (rare-earth) phosphors first used in the late 1960's are now almost universally employed; the original NTSC-prescribed chromaticities are difficult to attain and have been redefined in anticipation of HDTV (Zavada, 1988).

brightness axis from hue information. Most of the underlying models have been reviewed (Foley et al., 1990); an experimental comparison of the utility of the spaces based upon timing of interactive tasks has also been conducted (Schwarz, Cowan and Beatty, 1987).

### YIQ Space

The $YIQ$ space (McIlwain and Dean, 1956) used in North American television broadcasting records monochromatic *luminance* information along the $Y$ axis and hue *chrominance* on the in-phase and quadrature channels. Discussion and computational aids for guaranteeing that specifications lie within gamut, plus tabulations of conversion matrices for HDTV-defined phosphors appear elsewhere (Martindale and Paeth, 1991). The $Y$ channel is the luminous efficiency computed from the chromaticities of the NTSC-defined phosphor set. An orthogonal change of basis of two colour-difference axes establishes the mapping from $YIQ$ onto $RGB$ and vice versa. The choice is based on pioneering empirical studies on the eye's sensitivity to quantization effects of various colours (Stenger, 1977), allowing an encoding of the composite signal in decreasing bandwidths of 3.5MHz, 1.5MHz and 0.5MHz. This technique has been generalized to the encoding of chromaticity classes (In Der Smitten, 1974). Analogous methods of reflectance spaces whose coordinates have variable bit precisions are discussed in §8.2.2. A comparison of $YIQ$ versus $RGB$ in graphics applications has also been conducted (Smith, 1979).

### HSV and HLS Space

The $HSV$ (Smith, 1976) and $HLS$ formulations (ACM, 1979) commonly used in computer graphics also define an achromatic axis. (There is strong evidence that

colour perception at higher levels of cognition likewise make this separation (Cowan and Ware, 1985).) In both cases $V = R + G + B$ (the $HLS$ model derives lightness $L$ from $V$). Both models then depart from colour cubes and must instead index a bi-pyramidal model which approximates a bi-conical one. Cross-sections of the solid with a plane normal to its main (luminance) axis form hexagons, which encode chromatic information in two remaining degrees of freedom. The *hue angle* $H$ indicates orientation relative to the grey axis and saturation $S$ expresses the perpendicular distance to this axis. Because of the discrete turns occurring along the hexagon (belt projection), the hue angle is not a circular trigonometric function but instead an index indicating one of six colour sectors, with a fractional value further locating it along a sector's edge. Recording hue in non-angular fashion leads to computational speed-up but can burden the computation of the inverse which maps back onto RGB coordinates, though a fast algorithm has been presented (Fishkin, 1990). Alternately, a fast approximation to the arctangent may be employed (Capelli, 1991). A reflectance analog appears in Figure 5.2, by which point the linear reflectance spaces (Chapter 4) have been made symmetric (Chapter 5), engenering reflectance solids exhibiting pronounced cubic and bi-conical symmetry.

### RGB Space

Finally, the $RGB$ of phosphor values may be used as a basis *per se*. (This basis is all too often employd without any knowledge of the chromaticities of the device on which it is used.) Regarding these as orthogonal axes and setting the conditions of maximum output at $R = G = B = 1$ defines the conventional RGB *colour cube*. An geometric colour algorithm (Heckbert, 1982) computes a representative palette of pixel (colours) by rectilinear bisection of this cube based on histogram analysis of the populations of pixels appearing in a common scene. The hue hexagon normal to

the body diagonal can be augmented by three other interlocking hexagons normal to the remaining body diagonals (Holden, 1971), forming a useful partitioning of colour space by *a priori* means. This cubeoctohedran model is derived by both geometric and algebraic means in (Paeth, 1990a) and extended in (Paeth, 1991). Although the method is presented as a means of encoding RGB pixels on frame buffers of limited precision, it also serves as a geometric examination of the interior of a colour solid; most treatments (including this) study merely the solid's surface (extreme points). Other geometric partitionings of RGB and related additive colour space are known (Turkowski, 1986).

## 3.3 Summary and Conclusions

Methods of linear algebra and their related geometry have been developed which describe the additive colour of lights. The conditions of colour matching were established. These allowed the computation of (additive) metamers by employing quadratic forms based on generation and detection matrices. The *colour space* and related *colour solid* and *gamut* were given a geometrical footing to be used in the following chapters.

# Chapter 4

# Linear Reflectance Space

This chapter explores reflective colour spaces using analytical and geometric methods. Two prefacing sections follow. The first reviews and reifies the concepts of linear reflectance space through extension of analogous additive methods. The reflectance solid is then introduced and its geometric construction derived. Reflectance solids are created from empirical data using conventional regression techniques and then examined. Their geometric asymmetry suggests modifications to their derivation, described in the next chapter.

## 4.1 Algebraic Reflectance Space

Reflectance spaces were introduced in §2.3 in which the equation for surface reflectances [2.26]

$$\rho(\lambda) = x_1\beta_1(\lambda) + \ldots x_m\beta_m(\lambda) \qquad [4.1]$$

and the equation defining the conditions of physically valid reflectance [2.3]

$$0 \leq \rho(\lambda) \leq 1 \qquad [4.2]$$

were given. Note that [4.1] does not describe any physical process and that [4.2] introduces physical constraints.

The continuous reflectance function $\rho(\lambda)$ fully discloses[1] the surface colour of a physical object by recording its reflectivity as a function of electromagnetic wavelength. Reflectance functions are often treated as a set of related elements, which may be indexed. For instance, a *print gamut* $\rho_1(\lambda) \ldots \rho_{4096}(\lambda)$ describes the reflectances of a collection of coloured surfaces, as produced systematically on a digital printer.

Similarly, $\beta_i(\lambda)$ describes an $i$th basis element. Basis elements may be manipulated in a fashion similar to any reflectance function $\rho(\lambda)$; the distinguishing name $\beta$ suggests "basis". An element $\beta(\lambda)$ is a mathematically abstract reflectance functions not subject to the conditions of [4.2], since [4.1] need not describe any physical process. Indeed, conditions of orthogonality may create a basis whose elements have negative regions, thus not directly approximating any physical surfaces within the set of reflectances, but instead recording variances among members within that set.

In general, reflectances bases are *orthogonal.* This condition maximizes least-squares goodness of fit, is a consequence of many systematic processes of basis generation (§4.1.10) and facilitates the geometric modeling and algebraic manipulation of reflectance solids and their spaces. This choice follows historical trends in the modeling of reflectance spectra, beginning with Wyszecki (quoted in §6.1.1). Extensions which remove this constraint are described in §8.3.2.

The following subsections describe the nature of reflectance spaces and their bases with exposition analogous to sections §3.1.3 through §3.1.8. Because the

[1]This is subject to the conditions of viewer and sampling geometry, plus the effects of surface fluorescence. See §8.4.3.

linear superposition of basis elements has no physical analog, a means of basis construction is also described.

### 4.1.1 Reflectance Bases

Unlike additive spaces, these basis elements do not participate directly in the definition of a physical operation, such as detection or generation. Instead, the elements define a colour space which encodes information about (continuous or discrete) spectral power distributions using finite linear systems. This is not a liability: the quadratic forms involving both generation and detection matrices may now use a single basis matrix in order to map interchangeably between reflectance coordinates and reflectance functions.

Modeling a very small set of reflectances is a simple task: the basis set is merely the set of reflectances potentially encountered. Since printer gamuts are typically quite large this method is not generally practical. Instead, an orthogonal basis of reduced dimension may be employed, whose elements are abstract reflectance functions.

Likewise, the equations analogous to those of perceptual metamerism [3.21] provide for *reflectance matching* by establishing points of coincidence in an $m$-dimensional reflectance space. In particular, the reflectance match occurs entirely within a space of reflection, without any reference to a source of illumination.

### 4.1.2 Notations

The reflectance function $\rho(\lambda)$ is discretized when sampled at $n$ points, where $n$ is a (large) value fixed by the limits of spectroradiometry. This creates a *reflectance vec-*

*tor* $\mathbf{r}$. The $n$-dimensional reflectance function $\mathbf{r}$ is considered an exact representation of any continuous reflectance made discrete. The basis functions $\beta_1(\lambda) \ldots \beta_n(\lambda)$ are likewise discretized, and become the basis vectors $\mathbf{b}_1 \ldots \mathbf{b}_n$. In practice $\mathbf{b}$ conforms (by row) to $\mathbf{r}$ as the former are typically derived mathematically from the latter (sets of) reflectance data. Both are indexed by a common underlying function of wavelength $\lambda$, creating a row index. This is most often a linear mapping.

In like manner, the subset of $m$ independent reflectance functions $\mathbf{b}_1 \ldots \mathbf{b}_m$ define an $m$-dimensional *reflectance basis* which likewise spans a reflectance (sub)space. These are adjoined to form an $n \times m$ rectangular matrix $\mathbf{B}$, whose $i$th column is $\mathbf{b}_i$. Equation [4.1] may then be expressed using the simple matrix notation

$$\mathbf{r} = \mathbf{B}\mathbf{x} \qquad \text{[reflectance coordinates onto vectors]} \qquad [4.3]$$

in which the column vector $\mathbf{x}$ are the *reflectance coordinates* of a point lying in $m$-dimensional reflectance space. The notation emphasizes that $\mathbf{r}$ is a valid reflectance: it satisfies the law [4.2], while any column vector $\mathbf{b}$ need not.

Whereas additive spaces can be usefully restricted to three dimensions because of properties of human vision (the XYZ tristimulus space), no analogous principle can be employed with subtractive models. Because the representation of reflective surfaces must remain independent of both observer and source, spaces of (arbitrarily) high dimension are potentially required. This suggests that the dimension $m$ must approach the dimension of discretization $n$ in order to form an orthogonal basis set of sufficient dimension to provide accurate approximations for any spectra under study. In fact, practical models with $m \ll n$ are acheivable. Spectra not defined by any single surface (as when current radiosity methods take spectra into account) are not a central concern. Methods for the determination of columns of a rectangular matrix $\mathbf{B}$ given a set of surface reflectances is reviewed in §4.1.11.

### 4.1.3 Reflectance Coordinates and Projection

The column vector $\mathbf{x}^{(\mathrm{B})} = (x_1, \ldots, x_m)^{\mathrm{T}}$ above defines a set of *reflectance coordinates* for the reflectance vector $\mathbf{r}$ lying within the reflectance space defined by $\mathbf{B}$. The superscript notation of $\mathbf{x}^{(\mathrm{B})}$ makes explicit the membership of $\mathbf{x}$ within the subspace spanned by $\mathbf{B}$ and suggests its defining matrix multiplication in [4.3] above. This equation may be considered a mapping which carries $m$-vector reflectance coordinates onto reflectance vectors. The mapping is lossless (regardless of dimension $m$) because the $n$-vector $\mathbf{r}$ lies within $\mathbf{B}$'s span, by definition. This mapping is analogous to the generation equation of light presented in matrix form in [3.6].

The reverse[2] mapping carries a reflectance vector onto a set of reflectance coordinates, an analog to the detection equation [3.7]. The mapping potentially introduces loss: an arbitrary reflectance $\mathbf{r}$ need not lie wholly within the reflectance space spanned by fixed $\mathbf{B}$. With $\mathbf{B}$ orthogonal, the reflectance function $\mathbf{r}$ has a best approximation (nearest point) $\tilde{\mathbf{r}}^{(\mathrm{B})}$ found by (orthogonal) projection. After (Apostol, 1969b), the approximation's $i$th reflectance coordinate $x_i$ is computed as

$$x_i = \frac{\mathbf{b}_i \cdot \mathbf{r}}{\mathbf{b}_i \cdot \mathbf{b}_i}. \qquad [4.4]$$

Enforcing orthonormality among $\mathbf{B}$'s column vectors clears the denominator. The reflectance coordinates may be found *en masse* by taking advantage of matrix multiplication again. Here,

$$\tilde{\mathbf{x}}^{(\mathrm{B})} = \mathbf{B}^{\mathrm{T}}\mathbf{r} \qquad \text{[reflectance vectors to coordinates]} \qquad [4.5]$$

locates the perpendicular point of projection of reflectance vector $\mathbf{r}$ onto reflectance coordinates $\tilde{\mathbf{x}}$. That is, $\mathbf{B}$ provides a rectilinear coordinate frame (under some rigid

[2] This can be considered an inverse only when $m = n$.

rotation) for the reflectance space as a consequence of column orthogonality. The non-physical reflectance elements serve an analogous role with the abstract detector responsivities used to establish metameric matches in §3.1.7. When $m < n$, a suitably chosen set of basis elements is desirable. Dimensional considerations and basis construction may now be considered.

### 4.1.4 The $n$-Dimensional Reflectance Model

When $m$ is increased to $n$, $\mathbf{B}$ becomes a square matrix, providing sufficient degrees of freedom for reflectance coordinates to define any arbitrary reflectance exactly. In this case the basis most naturally employed for this parent $n$-dimensional space are the tine elements, as in Chapter 2. These orthogonal functions are diagonal, making $\mathbf{B}$ the $n \times n$ identity. Substitution of $\mathbf{B} = \mathbf{I}$ into [4.3] shows that

$$\mathbf{r} = \mathbf{Ix} \qquad \text{[reflectance coordinates and vectors agree in n-space]} \qquad [4.6]$$

holds exactly. Row equality indicates that each reflectance coordinate controls a single component of the reflectance function. Considering the mapping in terms of dimensional analysis (Taylor, 1974a), $\mathbf{B}$ maps a vector of dimensionless reflectance coordinates onto a reflectance function. The single symbols of compact matrix notation do not easily accommodate the introduction (as with scalar quantities) of physical units which distinguish between physical functions and their mathematical abstraction. Fortunately, the elements of $\mathbf{r}$ are ratios and hence are dimensionless.

Procedurally, the introduction of diagonal matrices casts new light upon the traditional algorithms used to compute reflective composition, as in the radiosity-based rendering packages. The term-by-term "cross-multiply" used to form the system product appearing in [2.23] may be presented in pseudo-Pascal:

```
FUNCTION SpecMult(v1, v2: REAL[1..n]; n: INT): REAL[1..n]
        INT i;
        FOR i := 1 to n DO
                SpecMult[i] = v1[i] * v2[i];
                OD
        RETURN(SpecMult);
        END.
```

Here the two reflectance vectors $\mathbf{v}1$ and $\mathbf{v}2$ are composed, returning the system reflectance function $\mathbf{v}1 \odot \mathbf{v}2$. The operation is expressible in linear algebraic terms as the product of two diagonal matrices,

$$\begin{bmatrix} x_{1,1}y_{1,1} & & 0 \\ & \ddots & \\ 0 & & x_{n,n}y_{n,n} \end{bmatrix} = \begin{bmatrix} x_{1,1} & & 0 \\ & \ddots & \\ 0 & & x_{n,n} \end{bmatrix} \times \begin{bmatrix} y_{1,1} & & 0 \\ & \ddots & \\ 0 & & y_{n,n} \end{bmatrix}. \qquad [4.7]$$

Thus, the traditional first-principles algorithm for composition of reflectance functions is in fact a matrix product operating upon reflectance coordinates. These diagonal matrices are represented by compact diagonal vectors. In this sense, it is a multiplication of sparse matrices. More generally, the mapping of reflectance coordinates need not use strictly diagonal forms.

### 4.1.5 The $m$-Dimensional Reflectance Model

When a reflectance space has dimension $m$ with $m < n$, a discrete model of reduced precision is created. Useful reflectance models seek to minimize the dimension $m$ while maintaining perceptual authenticity. Any $m$ dimensional reflectance space is a proper subset of the parent $n$ space, because any reflectance expressible in the $m$-space may be expressed losslessly as a discrete reflectance vector in $n$-space. However, two dissimilar subsets are possible: two $n \times m$ bases $\mathbf{B}$ and $\mathbf{B}'$ need not span the same subspace.

A reflectance space of reduced dimension $m < n$ must approximate a given reflectance function $\mathbf{r}$ using a vector $\tilde{\mathbf{r}}$. (As in §3.1.6, the " ˜ " signifies "approximating"). Unlike the case of illumination space, neither a source nor a set of (perceptual) detectors may be assumed. Instead, the functions must lie in close agreement across their domain of representation. Otherwise, a source whose spectral power distribution is chosen by an adversary[3] could emphasis their disparity at the wavelength of poorest fit, providing a counter-example to the conditions for a match. Here "close agreement" minimizes the distance function $\Delta(\mathbf{r}, \tilde{\mathbf{r}})$ which may be calculated using the inner product:

$$\Delta^2(\mathbf{r}, \tilde{\mathbf{r}}) = (\mathbf{r} - \tilde{\mathbf{r}}) \cdot (\mathbf{r} - \tilde{\mathbf{r}}) = ||\mathbf{r} - \tilde{\mathbf{r}}||^2. \qquad [4.8]$$

Using the Euclidean norm, this gives

$$\Delta^2(\mathbf{r}, \tilde{\mathbf{r}}) = \sum_{i=1}^{n} (\mathbf{r}_i^2 - \tilde{\mathbf{r}}_i^2) \qquad [4.9]$$

so that minimizing this distance becomes a problem in least-squares minimization (Lawson and Hanson, 1974).

### 4.1.6 Reflectance Coordinate Mapping

Reflectance coordinates may be mapped onto other reflectance coordinates, a primary concern. The most basic case considers the mapping to and from reflectance functions within the common space spanned by $\mathbf{B}$. Substitution of [4.3] into [4.5] gives:

$$\tilde{\mathbf{x}}^{(\mathrm{B})} = \mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{x}). \qquad \text{[reflectance coordinates onto self]} \qquad [4.10]$$

The $m \times m$ matrix $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ is the identity. Its appearance indicates that the reflectance coordinates $\mathbf{x}$ in $m$ space may be mapped losslessly onto the parent $n$ space and

[3] A hypothetical *nemesis* is often invoked in the verification of algorithms (Rawlins, 1992).

back again onto $\tilde{\mathbf{x}}^{(\mathrm{B})}$ with $\tilde{\mathbf{x}}^{(\mathrm{B})} = \mathbf{x}$. This is expected, since $\mathbf{x}$ defines a point lying in the reflectance subspace spanned by $\mathbf{B}$. This confirms that the dimensional increase of vector $\mathbf{x}$ onto its discrete reflectance is lossless. When the bases are dissimilar, more general cases are possible.

### 4.1.7 Reflectance Function Mapping

An arbitrary reflectance vector $\mathbf{r}$ can be approximated by a reflectance vector $\tilde{\mathbf{r}}^{(\mathrm{B})}$ lying within the reflectance space defined by $\mathbf{B}$. This is done by mapping $\mathbf{r}$ onto reflectance coordinates $\mathbf{x}^{(\mathrm{B})}$ and back onto a reflectance vector $\mathbf{r}^{(\mathrm{B})}$. Reversing the order of substitution of [4.3] and [4.5] gives

$$\tilde{\mathbf{r}}^{(\mathrm{B})} = \mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{r}) \qquad [4.11]$$

which may be rewritten as

$$\tilde{\mathbf{r}}^{(\mathrm{B})} = (\mathbf{B}\mathbf{B}^{\mathrm{T}})\mathbf{r} \qquad \text{[reflectance vectors to self]} \qquad [4.12]$$

since matrix multiplication is associative. This removes partial products expressing reflectance coordinates. This equation provides a *reflectance match.*

This approximating transformation defined by $(\mathbf{B}\mathbf{B}^{\mathrm{T}})$ is the $n \times n$ square *covariance* matrix of $\mathbf{B}$. Being both real and symmetric, it has real, orthogonal unit eigenvectors (Apostol, 1969c) which form a basis spanning $\mathbf{B}$. Orthogonality among the $m$ columns of $\mathbf{B}$ implies a covariance having unit eigenvalues with multiplicity $m$. The remaining $n - m$ eigenvalues are zero, since $\mathbf{B}$ has rank of only $m$.

The matrix $\mathbf{B}^{\mathrm{T}}$ maps onto reflectance coordinates; $\mathbf{B}$ does the reverse, analogous to the (dissimilar) detection $\mathbf{D}$ and generation $\mathbf{G}$ matrices of [3.11]. Substitution as by [3.16] forms the Moore-Penrose pseudo-inverse $\mathbf{P}+ = \mathbf{B}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}$ (Albert,

1972), (Golub and Kahan, 1965) whose transpose serves as a multiplicative inverse. Here $\mathbf{B}^{\mathrm{T}}$ is the identity by column orthogonality, leaving $\mathbf{P}^{+} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$ which is singular, as $\mathbf{P}^{-1}$ remains undefined. (Other contemporary applications of the pseudo-inverse to reflective colorimetry exist (Hurlbert and Poggio, 1988).) These covariance forms are important to dimensional deflation and will be taken up later.

### 4.1.8 Reflectance Matching

The reflectance coordinates of both the original reflectance $\mathbf{r}$ and the approximate $\tilde{\mathbf{r}}^{(\mathrm{B})}$ match. This is easily verified by premultiplication of each by $\mathbf{B}^{\mathrm{T}}$ in accordance with [4.5]:

$$\begin{aligned} \mathbf{B}^{\mathrm{T}}\tilde{\mathbf{r}}^{(\mathrm{B})} &= \mathbf{B}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}})\mathbf{r} \\ \mathbf{B}^{\mathrm{T}}\tilde{\mathbf{r}}^{(\mathrm{B})} &= (\mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{B}^{\mathrm{T}}\mathbf{r} \\ \mathbf{B}^{\mathrm{T}}\tilde{\mathbf{r}}^{(\mathrm{B})} &= \mathbf{B}^{\mathrm{T}}\mathbf{r} \end{aligned} \qquad [4.13]$$

with the $m \times m$ identity removed from the right-hand side.

By analogy with [3.20],

$$\mathbf{B}^{\mathrm{T}}(\tilde{\mathbf{r}}^{(\mathrm{B})} - \mathbf{r}) = 0, \qquad [4.14]$$

so the approximation is exact within the subspace defined by $\mathbf{B}^{\mathrm{T}}$. Thus, in any reflectance match, *the reflectance coordinates of the original and the approximation are identical.* Note that reflectance matching makes no reference to a source of illumination. Equation [4.13] defines an equivalence within a subspace of projection, in direct analogy to the detection equation [3.18] defining the conditions for metameric matching of light. Thus, $\tilde{\mathbf{r}}^{(\mathrm{b})}$ and $\mathbf{r}$ may be considered "reflectance metamers". Because the term "metamer" has a well-established meaning in the context of the matching of coloured lights, the term *reflectance match* is chosen preferentially.

Reflectance matching does not take place within the subspace of detection, such as the tristimulus space related to the cone responsivities of a standard observer. Instead, the match holds strictly in reference to the basis **B**. By reexpressing both **r** and $\tilde{\mathbf{r}}$ in the parent $n$ space, the goodness of fit may employ the geometric distance between this two points, after [4.9], a measure of variance. Methods of orthogonal basis construction which minimize this loss over a given set of reflectance functions (points) are standard regression techniques, described in §4.1.11 below and generalized in the extended study of deflation (Chapter 6). The nature of the null space of dimension $n - m$ lost in $\tilde{\mathbf{r}}$'s space of lower dimension is treated in the study of inflation (Chapter 7).

### 4.1.9 Deflation and Inflation

Any reflectance basis of $n$ elements spans a common parent space, thereby providing a point of common ascent for dimensional mapping. Representation at this level is lossless as suggested by [4.3]. For example, the locus of 4096 points lying in any reflectance $n$-space each describe their corresponding reflectance functions exactly.

Reflectance functions lying in spaces of dissimilar basis or dimension may be mapped interchangeably using the mapping equations listed above. In general, the mapping between reflectance spaces involves the dimensional *inflation* which carries a representation onto the common $n$-space, followed by a subsequent *deflation* onto the new basis. This approach is analogous to the additive operations set forth in §3.1.4 and is suggested by the Colour Hierarchy depicted in §2.3.1. Since all reflectance spaces are linear, any inflation or deflation is a transformation representable by a rectangular matrix transformation **M** which maps the reflectance vector $\mathbf{x}^{(\mathrm{B})}$ onto $\mathbf{x}^{(\mathrm{B}')}$.

Still unknown is a means for establishing the orthogonal column vectors which define the basis **B** of any reflectance space. Two methods are presented below.

### 4.1.10 The Construction of Orthogonal Bases

The determination of $m$ elements $\mathbf{b}_1 \ldots \mathbf{b}_m$ which form the basis **B** is now described.

#### A Priori Methods

The orthogonal basis elements may be chosen independent of any reflectance data. The discrete representations of common orthogonal polynomials (Stahl and Totik, 1992) include sinusoids, Walsh functions or Haar transform elements. Related representation methods are the those of Fourier (Bracewell, 1978), Hadamard (Chihara, 1978) and the wavelet theory of increasing popularity (Daubechies, 1990), (Daubechies, 1992), (Strang, 1989). For instance, the *discrete cosine transform* (Ahmed, Natarajan and Rao, 1974) or *DCT* forms a set of real orthogonal functions $\mathbf{b}_n = \cos n\theta = \mathrm{Re}(e^{ni\theta})$, where $0 \leq \theta < 2\pi$ defines one period in the rows of $\mathbf{b}_j$. This discrete representation provides monotonic convergence should the input function be band limited (that is, have a power spectrum decreasing with frequency (Bracewell, 1978)).

Prior research describes reflectance models which employ Fourier bases, either resembling those above (Buchsbaum and Gottshalk, 1984) or in general (D'Zmura and Lennie, 1986)[4]. The decomposition of reflectances as a harmonic series is not unreasonable, as typical reflectance functions are both continuous and band-limited (Stiles, Wyszecki and Ohta, 1977). Representation of reflectances using other predefined orthogonal polynomials (Eves, 1987) has also been studied, including the

[4]Many of the results of that research are subsumed by the work appearing in §5.3.3.

Legendre polynomials (Novak and Shafer, 1990) who fit a basis having reduced dimension. In general, reflectance function may be represented with vanishingly small error with arbitrarily large choice of $m$, assuming the function contains only certain[5] singularities. In practice, the error converges to zero as an information-theoretic consequence of band limitation (Panter, 1965).

Polynomial curve fitting of reflectances was proposed very early on (Moon, 1945). Using such schemes, the polynomial representing a composition of reflectances as in [2.23] has a degree which is the sum of the degrees of the constituent parts. A closed system of description has been explored (Raso, 1990) wherein Chebyshev approximation is employed to refit the product polynomial with an approximating polynomial of original degree. This method supports operations within a linear colour space of fixed dimension (whose basis elements are monomials) but introduces loss of accuracy through the dimensional reduction present with each subsequent composition.

### A Posteriori Methods

Related thesis research (Maloney, 1985a), (Maloney, 1985b) appropriates traditional methods of regression analysis. Maloney's basis production is an *a posteriori* method: an algorithmic procedure constructs the basis after it is presented a set of typical spectral reflectance curves. By using the singular value decomposition or *SVD,* a superior method of basis determination can be created. Tight perceptual bounds may be placed when applied to the reflectances of natural surfaces (Krinov, 1947). Using the linear basis model appearing in [4.1], the number of elements $m \approx 6$ is typical.

[5] More precisely, if the function is Lesbesque integrable (Rudin, 1973), with the number of removable singularities $s$ potentially infinite but countable: $s \in \aleph_0$.

Although the SVD is not a colorimetric method *per se,* it provides a useful starting point in the construction of **B**. As shown later, constraints may be introduced which modify the nature of basis elements to account for physical properties of reflectances and provide other geometric advantages.

### 4.1.11 The Singular Value Decomposition (SVD)

This procedure is used to provide a basis of reduced dimension $m$ which provides least-squares goodness of fit over a collection of data having $n$ degrees of freedom (samples per datum). The method begins by computing the covariance of a set of data. The basic technique (Golub and Loan, 1989a) as well as its statistical applications (Higham and Stewart, 1987) appear in standard texts. The SVD enjoys wide-spread computational utility and underpins many software libraries (Anderson et al., 1991) because it operates freely upon singular matrices. Algorithmic implementations are available (Press et al., 1992), numerical stability and conditioning are well studied (Cook and Weisberg, 1982) together with perturbation analysis (Stewart, 1979), (Stewart, 1984).

The covariance matrix $\mathbf{C} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$ of any given $n \times m$ data matrix $\mathbf{D}$ is a symmetric matrix and hence has real eigenvalues and orthogonal eigenvectors. Moreover, it is positive semidefinite, so that $\mathbf{x}\mathbf{C}\mathbf{x}^{\mathrm{T}} \geq 0$ for all $\mathbf{x} \neq 0$. This holds $iff$ $\mathbf{C}$ contains non-negative eigenvalues (Leon, 1990b). (Such a matrix $\mathbf{C}$ can be reexpressed as the unique $n \times n$ product $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ by Cholesky factorization (Golub and Loan, 1989b).) Choosing as $\mathbf{x}$ the set of (unit) eigenvectors, then in light of orthogonality, the matrix may be expressed as the sum of the outer product of its $n$ orthogonal eigenvectors $\mathbf{v}_i$, scaled by their associated eigenvalues $\mu_i$:

$$\mathbf{C} = \mu_1 \mathbf{v}_1 \mathbf{v}_1^{\mathrm{T}} + \ldots + \mu_n \mathbf{v}_n \mathbf{v}_n^{\mathrm{T}}. \qquad [4.15]$$

Here the (columnar) eigenvectors $\mathbf{v}_1 \ldots \mathbf{v}_n$ are the *principal components* of the data set (Hotelling, 1933). Their associated non-negative eigenvalues represent the squared variance of that axis of representation (Wesolowsky, 1976). This is the singular value decomposition on $\mathbf{D}$.

Previously, the $n \times m$ orthogonal basis $\mathbf{B}$ produced a symmetric $n \times n$ covariance $\mathbf{C} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$ having reduced rank $m$. The loss in rank (the degree of singularity) is easily verified by finding the eigenpairs on $\mathbf{C}$: the $n$ eigenvalues may be arranged in a diagonal matrix in which the first $m$ are unity, the remaining $n - m$ zero. The corresponding initial $m$ principal components (eigenvectors) form an orthogonal set which fully span the data. The remaining $n - m$ elements define the null space of representation which may be removed. For a general data matrix $\mathbf{D}$, the eigenvalues are non-negative and dissimilar. They are customarily arranged in order of decreasing value. Retaining merely the first (last) $m$ associated eigenvectors creates a subspace of representation which minimizes (maximizes) the least squares variance of the data set.

As an example, consider the CRT model: a linear process having $m = 3$ independent controls and which produces a large set of $n$-dimensional spectral power distributions $\mathbf{s}$ which are characterized in the presence of perturbing noise. The SVD of the CRT data begins with the set of distributions of a large sampling $k$ of the *display gamut:* the set of all colours (distributions) when can be generated on the device under study. In this case the samples are recorded as the $n \times k$ data matrix $\mathbf{D}$, with $m < n < k$. Because of the additive nature of light rays, the rank of the $n \times n$ covariance matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ is also three. Here $m < n$ and the SVD recovers the $m$ degrees of freedom and indicates the "nearly null" space of noise by locating the trailing components whose associated eigenvalues are near zero. Thus, the SVD generalizes the concept of rank and singular value.

It is worth noting that the underlying spectral power distributions can only be recovered to within a linear change of basis. Moreover, SVD necessarily yields orthogonal set of components. There is a very high probability that the principle component (initial eigenvector) will be everywhere positive (due to the positive elements for all data presented to the algorithm). As a consequence, successive eigenvectors must then include negative components to enforce orthogonality. Thus, a basis derived by SVD rarely describes physical sources. Therefore, orthogonal systems (and the SVD) are generally inappropriate for modeling linear additive colour. Instead, a set of independent distributions characterizing the physical sources provides an adequate, physically-based model.

By analogy, the SVD may be applied to a *printer gamut* of normalized reflectances $\mathbf{r}_1 \ldots \mathbf{r}_k$ in which each $\mathbf{r}$ samples a representative printer colour. Here $k > n$ is typical. Augmentation forms $\mathbf{R}$, an $n \times k$ matrix of reflectance data which is a discrete sampling of the full gamut. The eigenpairs of the covariance $C = \mathbf{R}\mathbf{R}^{\mathrm{T}}$ are then computed. Because of non-linearity of the print transfer function and in particular, the failure of linearity under the composition of reflectance, the SVD will indicate a number of representational degrees of freedom greater than the number of independent processes. A typical printer has three or four dye layers. As mentioned in §4.10, $m \approx 6$ provides a tight fit on representing many naturally occurring reflectances; $m \approx 5$ proved more than adequate for the reflective media studied in Chapter 6.

### 4.1.12 The SVD: Practice and Extensions

Iterative methods which provide a partial decomposition in ranked order of dominant components are attractive, particularly with large data sets[6]. Typical requirements are data sets with $n = 512$ and $m < 10$. Iterative methods generate basis elements until suitable stopping conditions are met. Any SVD produces basis elements and companion diagonal elements by finding the eigenpairs associated with the data covariance matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$. As these matrices can be large, a custom library was created which provides on-demand evaluation of elements in $\mathbf{D}\mathbf{D}^{\mathrm{T}}$, described in Appendix E. Iterative solution employed the *Power method* [7] and subsequent *Weilandt deflation* based on the identity [4.15] to form a residual matrix. The techniques are describe in detail in Appendix A. This approach models state-of-the-art SVD solvers employing orthogonal block-matrix techniques (Scott, 1979), (Golub, Luk and Overton, 1981), (Overton, 1990) except a unit block size is used and QR (Gram-Schmidt) reorthogonalization is not required.

Typical stopping conditions occur when $m$ matches the target dimension set by the problem, or by a termination which occurs when an eigenvalue falls below unity or shows a pronounced decrease in value (Tatsuoka, 1988). A zero diagonal element indicates that the rank of the data set has been achieved: the given data is exactly representable using a subspace of less than $m$ elements. The SVD enjoys wide computational use as it is robust in the presence of singular data. In passing, none of the empirically derived data presented here has ever been singular, though elements quickly approach zero well before the final basis element.

In §5.1 modifications to the SVD suitable for the modeling of reflectance data are

[6] Initial copies of MATLAB could not provide full solutions of systems larger than about $200 \times 200$.

[7] Also known as the *Hotelling method* in older texts.

described and examined. Modifications and alternatives to the SVD in a more general context are described elsewhere (Hong and Pan, 1992), (Chan, 1986), (Hansen, Sekii and Shibahashi, 1991).

## 4.2 Reflectance Solids

A reflectance space is a linear space spanned by a basis of reflectance functions, after [4.1]. A point lying in this reflectance space defines a unique reflectance function $\rho(\lambda)$, computed as the linear combination of each basis element $\beta(\lambda)$ weighted by the components of the coordinates of the associated point in reflectance space. Because the $\beta(\lambda)$ are abstract and the space (domain of coordinates) infinite, the associated reflectance $\rho(\lambda)$ need not satisfy [4.2]. In this case, the associated point describes an *invalid reflectance.* The locus of all valid points defines the *reflectance solid,* which is unique to the basis defining its space of definition, making it a canonical object of study. The $n$-dimensional reflectance solid is written $\mathcal{R}^n$. The remainder of this chapter considers the geometric nature of reflectance solids.

In general, reflectance solids are convex polytopes. This may be demonstrated in the course of their construction. Because a reflectance solid describes the set of all physically valid reflectances, equation [4.1] is substituted into [4.2], giving

$$\mathcal{R}^n \quad \equiv \quad 0 \leq \sum x_i \beta_i(\lambda) \leq 1. \qquad [4.16]$$

This condition must hold for arbitrary $\lambda$. When the basis functions are discrete vectors $\mathbf{B}_{n \times m} = \mathbf{b}_1 \ldots \mathbf{b}_m$, the solid's volume may be created by a finite number of descriptive solid geometric operations on the given matrix.

The convex reflectance gamut of dimension $m$ is defined as the intersection of

the $2n$ linear inequalities

$$\mathcal{R}^n \equiv \bigcap_{l=1}^{n} \left[ 0 \leq \sum_{i=1}^{m} x_i \mathbf{B}_{l,i} \leq 1 \right]. \qquad [4.17]$$

in which $l$ serves as an indexing parameter analogous to $\lambda$, providing $n$ distinct wavelengths (the row number in the $n \times m$ basis matrix). These $m$ elements $\mathbf{b}_1 \ldots \mathbf{b}_m$ define a unique solid in $m$-space. The solid is the locus of all points $(x_1 \ldots x_m)$ which satisfy the simultaneous linear equations in [4.17], with a point lying on the solid's surface if at least one condition is satisfied by equality. Re-expressing $\mathbf{B}$ as column vectors and writing the interval test as two simultaneous inequalities gives

$$\begin{aligned} x_1\mathbf{b}_1(\lambda) + \ldots + x_m\mathbf{b}_m(\lambda) - 1 &\leq 0 \\ x_1\mathbf{b}_1(\lambda) + \ldots + x_m\mathbf{b}_m(\lambda) &\geq 0, \end{aligned} \qquad [4.18]$$

in which the first (second) equation defines the upper (lower) boundary of each halfspace associated with $\lambda$. In fact, each region is a *zone.* (The geometry and related terminology appear in §B.5 and §B.6.)

Although a wavelength indexes a unique zone the converse need not hold. Not all zones need contribute to the faceting of the solid. They may fully contain it or the facets[8] produced may be removed by a subsequent intersection. Thus, the solid may contain fewer than $2n$ facets.

### 4.2.1 Zone Placement within the Reflectance Solid

The reflectance solid is a convex polytope since it is created by repetitive halfplane partitioning of convex space, with each new partition preserving convexity. The planes occur in upper and lower pairs for each wavelength $l$. These share a common normal vector $\mathbf{v} = (\mathbf{b}_{l,1}, \ldots, \mathbf{b}_{l,m})$ and hence direction, but have distinct offsets,

[8]The geometric terminology and related operations are reviewed in Appendix B.

after the constants appearing in [4.18]. The paired planes are thus both parallel and distinct, verifying that they are zones. The $l$th zone in $m$ space is derived from the $m$ column entries appearing in $l$th row of $\mathbf{B}$. The zone is then associated with the sampling wavelength of this row. The presence of zones is a necessary condition in the definition of zonotopes, yet is not sufficient: the centre of each zone must also lie at the solid's centroid in order to provide a point of central symmetry (see Appendix B).

The thickness of any zone is the length of a segment perpendicular to the lower and upper planes. The lower planes all contain the origin, as substitution of $(0, \ldots, 0)$ always satisfies [4.18b]. Thickness is therefore the (perpendicular) origin distance to the upper plane. In (Hessian) normal form[9], the constant of offset gives the absolute length. Since the offset is initially one, the zone thickness is the reciprocal of the length of the unnormalized vector $\mathbf{v}$:

$$\text{thickness} = (\mathbf{v}\mathbf{v}^{\mathrm{T}})^{-1/2} = (\mathbf{b}_1^2[l] + \ldots + \mathbf{b}_m^2[l])^{-1/2}. \qquad [4.19]$$

Since $\mathbf{v}$ is the (wavelength) row of data matrix $\mathbf{D}$, the zone thicknesses are the reciprocal square-roots of the diagonals of the covariance matrix $\mathbf{D}^{\mathrm{T}}\mathbf{D}$ which forms the inner product of $\mathbf{D}$ by row. Thus, the zone thickness decreases as the distribution of associated sample points in the colour space lies closer to the hyperplane determined by regression.

Because the zone thicknesses are variable yet all meet and "pivot" about the origin, their centres need not be in common. This reflectance solid in general is a convex polytope not possessing additional geometric attributes.

[9] That is, after dividing through to form a unit vector of the components.

### 4.2.2 Canonical Nature

The reflectance solid provides a canonical form useful in recording the representational features of its related reflectance space. For any given $\mathbf{B}$ a unique solid may be constructed.

More important, when the columns of $\mathbf{B}$ are produced deterministically and in sequence, the solid is unique for the print gamut used to determine $\mathbf{B}$. This is the case with SVD. It is not a trivial consequence: should SVD (for instance) locate an eigenvalue[10] of multiplicity $j$, the ordering of $j$ related eigenvectors becomes undefined. However, any alternative set of eigenvectors sharing the multiple eigenvalue will be identical to within a linear change of basis. (Characterization of such multiplicity is closely related to the theory *Jordan blocks* (Golub and Loan, 1989c).) Since orthogonality must hold for repeated eigenvalues of any real symmetric matrix (Hamilton, 1989), the alternative set differs by a rigid rotation or reflection, thus preserving the reflectance solid's geometry. Conceptually, the solid is free to rotate about the $j$ multiple eigenvectors precisely because no principal axis dominates.

Additionally, any derivation may differ in the sign (direction) of any basis element $\mathbf{b}_i$. (In practice, this ambiguity manifested itself and was treated when custom SVD-based software was checked against otherwise identical results produced using MATLAB.) Complementation of basis vector $b_j$ alters the handedness of the coordinate frame by reflection, as seen by substitution into [4.18]:

$$x_1\mathbf{b}_1 + \ldots + x_j(-\mathbf{b}_j) + \ldots + x_m\mathbf{b}_m \;=\; x_1\mathbf{b}_1 + \ldots + (-x_j)\mathbf{b}_j + \ldots + x_m\mathbf{b}_m \quad [4.20]$$

thus reversing the indexing along that axis. Clearly, such mirrorings do not modify the surface geometry of the solid.

[10] This condition is not likely to arise in empirical practice.

The procedures applied here substitute the basis vector $-\mathbf{b}_j$ for any $\mathbf{b}_j$ to assure that the initial element of this eigenvector is non-negative. This forms a canonical solid for any printer gamut. In some cases, $\mathbf{b}_j[0] \approx 0$, as when the discrete reflectance elements show a tapering at the wavelength extremes. This suggests an alternate normalization procedure such as one enforcing $max(|b_j[1]| \ldots |b_J[n]|) > 0$, thereby placing the discrete reflectance vectors's greatest excursion on the positive side. (A canonical basis derivation has also been advocated (Parkkinen, Hallikainen and Jaaskelainen, 1989); the authors suggest similar methods in their concluding sections.) It is this canonical nature of the solid that makes geometric study of empirically determined reflectance spaces so attractive.

### 4.2.3 Empirically Produced Reflectance Solids

Colour solids were created based on SVD-derived vectors from empirically determined data sets. The first data set was a sampling of the gamut of Kodak Ektachrome 6184 roll film. The second was for Kodak Ektatherm SVC100 (thermal sublimation) paper. The spectroradiometric methods and practice were identical in both cases and have already been described (Paeth, 1989a), (Paeth, 1989b). The control files and in particular the sampling resolution was also identical, inviting direct comparison of results.

#### Transmissive Media

Kodak Ektachrome 6184 is a 35mm ISO[11] 100 positive (slide) film provided in cassettes or in bulk. The latter is not pre-aged for general consumer use is and

[11]Previously termed ASA.

Figure 4.1: SVD Transmission Basis (Ektachrome 6184)

available under the "P-professional" designation, to be refrigerated until immediately before use. Sixty-four frames, each of uniform colour (and from the same bulk roll) were imaged on a Dunn Instruments 631 film recorder. These provided a coarse $4 \times 4 \times 4$ Cartesian sampling of the printer's gamut. The 64 data points were reduced using SVD onto a set of reflectance vectors and associated eigenvalues. The singular value decomposition then produced a ranked set of the first eight orthogonal vectors. Because this is transmissive media, effects such as inter-layer scattering (Hanrahan and Krueger, 1993) and surface gloss are greatly reduced – a basis of three principal components provides a reasonable fit to the data. The components are plotted as piece-wise linear functions in Figure 4.1. The data shows an unexpected similarity to the first three Chebyshev polynomials (Apostol,

1969d),

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_2(x) &= 2x^2 - 1 \\ T_n(x) &= 2xT_{n-1}(x) - T_{n-2}(x) \\ &= \cos(n \cos^{-1}(x)), \end{aligned} \qquad [4.21]$$

within the domain $450nm \leq \lambda \leq 660$. Also note the negative excursions of the second and third components: they describe invalid reflectances. In particular, the chromaticity coordinates of $\mathbf{b}_2$ or $\mathbf{b}_3$ (under illumination such as by CIE Source E) are negative and hence lie outside the chromaticity diagram (a sufficient yet not necessary condition). Nontheless, a hue can be still be ascribed to them. By extrapolation, the reflectivity of the second and third elements describe "ultra-canary" yellow-green and "ultra-azure" blue-green. Ascribing colour to a non-physical chromatic axis is not uncommon (Buchsbaum and Gottshalk, 1984), (Stenger, 1977) (Parkkinen, Hallikainen and Jaaskelainen, 1989); see also §6.3.5. Generalization of these methods suggests the perceptually-based theory of *opponent colours* (Buchsbaum and Gottshalk, 1983), (Schwarz, Cowan and Beatty, 1987), taken up again in §7.4.6.

Conversely, the contribution of some element $\mathbf{b}_j$ may be negative, indicating the subtraction of a reflectance element $\beta_j(\lambda)$ from reflectance function $\rho(\lambda)$. That is, valid reflectances may have negative reflectance coordinates within their subtractive colour space, a situation not possible with the physical (additive) mixing of light.

The further study of transmissive media resumes in §8.1.

Figure 4.2: SVD Reflectance Basis (Ektatherm SVC100)

### Reflective Media

Reflective media in general yields samples less easily modeled. For comparison's sake, a $4 \times 4 \times 4$ Cartesian sampling of the Kodak SVC6500 thermal sublimation printer was conduced using identical control files and automated procedures. SVD was again applied, producing the plots appearing in Figure 4.2. A reflectance solid may now be constructed using the custom solid modeling package (see Appendix E). The surface colour of the reflectance solid at any point is rendered by computing the reflectance function at point and then converting into CIE XYZ chromaticity coordinates. Finally these are mapped onto RGB values based on the published phosphor chromaticities of the Silicon Graphics Iris, the hardware on which the package runs. A paper copy of the image appears in Figure 4.3. Because the reflective solid has no colour *per se,* its surface colour is computed under a hypothetical equipotential source, CIE Source E. (The tool allows the substitution of

Figure 4.3: SVD Reflectance Solid

alternate sources of illumination, see also Figure 6.9 and Appendix E.) The image is therefore *self-referential* – its spatial and radiometric features describe each other[12].

### 4.2.4 Coarse Faceting, Heuristics

A sparse sampling by wavelength reduces the number of zones and hence facets and vertices, greatly simplying computational complexity in rendering. (The software runs in time $O(z^2)$ with $z$ the number of zones to simplify the construction of colour solids in the fourth and higher dimension.) In Figure 4.3 discrete basis vectors were thinned by a $5:1$ ratio by dropping four of every five rows, thus retaining samples appearing at integral $5nm$ wavelengths. That is, the size of the reflectance basis $\mathbf{B}$ used to create the solid in 3-space is now $[75 \times 3]$ with row number $r$ representing a wavelength of $(365 + 5r)nm$, hence along the domain $[370nm..740nm]$.

[12]That is, when viewing the image on the CRT. A truly self-referential set of reflective media appears in [6.4.8].

The geometric solid created from the $z = 75$ zones (150 partitioning planes) contains $v = 60$ vertices, $f = 32$ faces and $e = 90$. The overall shape is preserved by the simplified, discrete model. This is significant: geometric feature thinning is in general a difficult operation (Matheiss and Rubin, 1980) which can have catastrophic results even upon self-consistent, robust polyhedral data structures such as the winged-edge model (Baumgart, 1974). Even when pathology such as dangling edges or planar dihedrals is accounted for, problems remain. In particular, the removal of an outlying vertex will collapse the convex hull of a polyhedron, possibly forming a polygon. For instance, the removal of the black vertex, as from the Figure 4.4, may well collapse the solid. (This can be of virtue, as described within §7.4.3.) The procedural difficulties in maintaining robust descriptions of polyhedra may be described in terms of graph-theoretic rules (Heisserman and Woodbury, 1993) or in terms of theorems (Sugihara, 1993).

Here, the simplified model is rederived from a thinned set of cutting planes. Because the reflectance functions defining the planes are typically $C^1$ continuous (a consequence of their derivation from data sets describing continuous reflectance functions), the rows in $\mathbf{B}$ which define the normal vectors and hence orientation of each zone are nearly identical. The reflectance solid is thus *sculpted* from convex space by a sequence of discrete partitionings which lie along a path of continuously changing orientation. Thinning the data increases the discrete step size and increases the coarseness of the facets. Whereas vertex culling is error prone, zone culling is not and enjoys advantages besides.

Note that the (near) continuous solid is fully encased by the coarse solid. Coarse solids provide a *bounding box,* making a useful containment heuristic. Such tests are generally valuable and are common to the practice of computer graphics (Glassner, 1990). Specifically, a fine-sampled solid having $z$ zones is always contained within

a coarse-sampled having $z' < z$ zones. Omitting partitions does not have the same potentially catastrophic consequences as does vertex culling. Instead, containment creates a nested sequence of approximating solids having zones $z'' \ldots < z \leq n$. Should an object not intersect the coarsest parent solid $z^{(\mathrm{m}')}$, neither can it intersect any descendent. This also explains the non-collapse of volume in the coarse parent. When thinned not by vertex culling but instead by zone culling through decreasing $z$, the limiting finite volume is the $m$-dimensional *parallelotope* having $2m$ sides and $m$ zones with $m < z' < \ldots < z \leq n$. Upon further thinning, a series of unbounded convex objects is created, not unlike the semi-open cubical illumination solid of §3.2.1. Ultimately, a non-solid having no zones defines the undifferentiated (convex) space which contains all solids.

### 4.2.5 Fine Faceting

Proceeding in the opposite direction, the number of zones may approach but cannot exceed the sampling dimension $n$. As $n$ in turn is increased to the power of the continuum, the continuous basis functions $\beta_1(\lambda) \ldots \beta_m(\lambda)$ "sculpt" reflectance solid in $m$-space, giving it continuous, curved surfaces. Although arbitrarily large sampling is impossible in practice, bases of continuous reflectance functions can be constructed from *a priori* methods. Moreover, continuous analogs to the SVD exist (Horn, 1950), allowing the derivation of a set of continuous orthogonal functions given a set of continuous distributions. Continuous reflectance functions can be derived with high accuracy for certain important surfaces, such as *dichroic* filters, or could be derived analytically from the physical properties of matter, much as the Planck's formula for black-body radiators in [2.5], or created by polynomial curve fitting (Raso and Fournier, 1991).

The solid in Figure 4.3 is already nearly "smooth" for any practical purpose – detail was not increased visibly for solids were created from more than three hundred wavelength samples (with $z \approx 300$). Still, certain features persist at the finer scale. These features are not artifacts of the sampling dimension $n$, but are essential features of the reflectance solid. As suggested by close examination of Figure 4.3, surface features of various dimension are present, including vertices, crease-like ridges and visible surface faceting. Their nature and existence is the subject of the following sections.

### 4.2.6 The Black Apex

All reflectance solids contain the origin: the vertex common to the lower defining plane of all $n$ zones. Since the multiplicity of planes $n$ is at least dimension of any space $m$, the feature is a vertex (see Appendix B) and hence an extreme point on the solid's surface. More important, $n > m$, so the vertex is *degenerate.* Being common to all zones of various orientation, the extremum forms an *apex,* or cusp. The black apex is present in all reflectance solids at a point of $C^1$ surface discontinuity. An illustration created from empirical data by retaining those facets which contain the black apex appear in Figure 4.4.

Algebraically, the point locates the reflectance function $\rho(\lambda) = 0$ with associated reflectance coordinates $\mathbf{x} = 0$. This non-reflector associated with the apex may be properly termed "black" in the absence of any external source of illumination. (Note that this term often cannot be defined, see [3.27].) The black reflectance lies at the boundary of the interval of valid reflectance for all discrete or continuous wavelengths $l$ or $\lambda$. It is significant that this reflectance function exists independent of the choice of basis elements $\beta(\lambda)$. A very important virtue may now be stated.

Figure 4.4: Facets Common to the Black Apex

*All reflectance spaces contain a reflectance solid which possesses a black apex.*

### 4.2.7 The Best White Apex

A conjugate white apex does not necessarily exist for a given basis. That is, the ideal reflectance $\rho(\lambda) = 1$ and hence any equipotential grey reflectance $\rho(\lambda) = c$ (a constant) derived by linear scaling need not lie in the $m$-dimensional reflectance space. Generally, the value (ordinate) of the reflectance function may be constrained at most $m$ selected values, as the associated coordinates $\mathbf{x} = (x_1, \ldots x_m)$ provide $m$ degrees of freedom. An exact representation implies that the discrete reflectance vector $\mathbf{r} = (1, \ldots, 1)^{\mathrm{T}}$, can be formed as a linear combination of the $m$ reflectance elements $\mathbf{b}_1 \ldots \mathbf{b}_m$. That is, the body diagonal of the parent hypercube in $n$-space must be preserved when an $m$-dimension subspace of deflation is created. Note that

most *a priori* polynomial orthogonalization techniques provide the unit monomial $u_0(x) = 1$ as the first member of their series and therefore satisfy this condition for any dimension $m$.

Given an arbitrary reflectance basis $\mathbf{B}$, the metameric white point of the reflective surface $\mathbf{r} = (1, \ldots, 1)^{\mathrm{T}}$ may be found using [4.5]. Here, the vector product sums the elements in each $i$th column of $\mathbf{B}$ to determine the reflectance component $\mathbf{x}_i$. Application of [4.11] then finds the approximating reflectance. Here the operation may be conducted along a continuous domain. Given $\rho(\lambda) = 1$, the coordinates used to provide a white reflectance match may be computed as

$$\rho(\lambda) = \sum_{i=1}^{m} \left[ \int \beta_i(\lambda)\ d\lambda \right] \beta_i(\lambda). \qquad [4.22]$$

For a basis of reflectance functions all having have unit area (such as the diagonal tine basis), the coordinates of the white point lie at $\mathbf{x} = (1, \ldots, 1)^{\mathrm{T}}$. For empirically-derived reflectance functions the areas under each reflectance are arbitrary. Summing the rows of all reflectance elements produces $[8.4920,\ -1.6199,\ -0.1996 \cdots]$ with all trailing elements (elided) having a value less than $|.005|$. These are the reflectance coordinates in $m$ space, here with $m = 64$. Retaining the first three entries forms a best approximation for the space $m = 3$. The approximating reflectance function is then produced by summing each basis element in relative contribution. The results are plotted in Figure 4.5.

Unfortunately, this least-squares approximation straddles the line $y = 1$, thus failing to satisfy the physical condition [4.2]. Thus, portions of the projected gamut may like outside of the reflectance solid, making it invalid. This situation must be remedied if reflectance gamuts are to be of practical value. That is, a *best white* must be found which both lies in gamut and which maximizes the value $\rho(\lambda)$ everywhere while leaving this function relatively flat (nearly constant, thus

Figure 4.5: SVD Solid White Point

showing little spectral bias). One method of correction scales the reflectance as $\alpha\tilde{\rho}^{(\text{white})}(\lambda)$ until it entirely within bounds. Procedurally, the wavelength of peak excursion $\lambda_{max}$ is located and the reciprocal of the reflectance at that index used as a scale $\alpha = [\rho^{(\text{white})}(\lambda_{max})]^{-1}$, thus adjusting the maximum value of the scaled reflectance to unity at $\lambda_{max}$. For the Kodak SVD-derived basis these occurred at $\lambda_{470nm} = .9270$ and $\lambda_{740nm} = 1.1345$, hence $\alpha = .8814$.

The reflectance coordinates $\mathbf{x}'$ corresponding to the particular solution are easily computed as $\mathbf{x}' = \alpha\mathbf{x}$ by virtue of linearity. The computational advantages of this approach (Martindale and Paeth, 1991) has been exploited exploited by the author to relocate "NTSC hot" pixels out of out of broadcast gamut (Smith, 1978b) by virtue of the linear ($3 \times 3$ matrix) transformation mapping RGB values onto YIQ values. (See also §3.2.5 and Appendix G.) The method works identically within both additive and reflective linear colour space. A possible drawback to this approach is that a particularly large peak at $\lambda_{max}$ may lower the out-of-gamut curve $\tilde{\rho}^{(\text{white})}(\lambda)$ unnecessarily, as only a single degree of freedom $\alpha$ is considered. This is clear from the figure: the maximum value occurs at the extreme end of the range, depressing the reflectivities throughout much of the visible spectrum.

The best white point can be also found by geometric methods. First, the point

must lie at the limit of the condition of valid reflectance, hence upon the surface of the reflectance solid. A point of closest approach is therefore sought. The best approximation may be determined by geometric means. The point of closest approach within the reflectance solid is found. For spaces in which the (weighted) Euclidean distance gives a measure of best fit, the solution is exact. By arguments of convexity, this point will lie either at the foot of perpendicular projection upon a facet or at that vertex of closest approach to the external white point (indicating that the relative minimum does not occur within the facet but at its boundary). The facet of closest proximity may be identified using a "brute-force" linear $O(z)$ search against the $z$ planes bounding the upper portion of each $z$ zones. (If efficiency is desired, a binary space partitioning of the solid make possible sublinear searches running in time $O(\log z)$.)

The nearest face is bounded by at least $m$ vertices. Each vertex is in turn associated with the (upper) boundary of a selected zone, indexed by wavelength. Thus, all points along the facet are in correspondence with the set of reflectance functions having a maximum at $y = 1.0$ for at least $m$ limiting wavelengths. Most often $m$ is also the upper bound, unless degenerate vertices exist. Local motion along the nearest facet changes the profile of the reflectance function while maintaining the functions value at unity for the set of constrained wavelengths. This solution allows the secondary maxima to be increased, tightening the overall goodness of fit.

For the empirical data presented above, the set of twenty-eight closest vertices form a common facet, accounting for the solid's snubbed appearance. Each vertex is also common to a facet containing the black apex. Thus, the reflectance function at any vertex in the set has both a maximum at $y = 1$ and a minimum and $y = 0$, none are good candidates for nearest white. Instead, a proper linear combination must be found to define a point lying in their common plane. This was not done.

In general, the surface of the solid defines points whose reflectances lie at the limit of validity for one given wavelength. Features of reduced dimension, such as edges or vertices show an increase in the number of limiting wavelengths. This is treated in the following sections.

### 4.2.8 The Surface and Its Reflectance Functions

The surface of the reflectance solid defines all reflectance functions which satisfy the limits of the physical condition [2.3] for at least one point. That is, there exists at least one wavelength $\lambda'$ for which either $\rho(\lambda') = 1$ or $\rho(\lambda') = 0$. The wavelength indexes a zone, the higher (lower) limit indicates an top (bottom) zonal plane. Reflectance functions exhibiting a multiplicity of limiting extrema are therefore common to multiple zones. As the zone multiplicity ranges from 1 to $m$ (the dimension of the reflectance space), the dimensionality of the related surface feature decreases from $m - 1$ to zero. For instance, the surface, edge and vertex points of a 3D reflectance solid are common to one, two or three respective defining zones. (A degenerate vertex, such as the black apex, is common to more than three.) The reflectance functions of a range of points lying along an edge or face may be easily determined by linear interpolation of the reflectance functions of the vertices which bound and define that feature. For this reason, the features most often studied are the vertices of the reflectance solid, which fully define the entire polytope as its convex hull.

Certain vertices show "mixed" upper and lower limits because they are common to the intersection of upper and lower planes. Such vertices often lie along a *crease* (dihedral line), that is, upon to dissimilar planes, one containing the black apex, the other the white. These functions are of particular importance as they define

simultaneous minimum and maximum ranging reflectances. They are described again in §4.2.10 and §7.5.

### 4.2.9 Geometric Artifacts of the Surface

When the sampling dimension $n$ is increased the solid becomes increasingly sculpted as the number of facets and vertices increases. Often, the new vertices do not reveal additional extrema distinct from lower-dimensional counterparts. Instead, the dihedral angle between adjacent facets moves to 180° and the surface is smooth. The new extrema (e.g., an edge) appears along the surface of a face and defines the zone of transition where all associated reflectance functions have a maximum (or minimum) both at the original wavelength and at the immediately adjacent wavelength introduced by the increase in $n$.

Algebraically, the extrema of any continuous reflectance functions occur at a point of zero slope, so that

$$\rho'(\lambda) = \frac{\rho(\lambda + \epsilon) - \rho(\lambda)}{\epsilon} \qquad [4.23]$$

Since the reflectance functions $\mathbf{r}[l]$ are discrete, then for some integral wavelength index $l$ there is a point of no slope under a forward difference equation:

$$0 = \frac{(\mathbf{r}_l - \mathbf{r}_{l+1})}{(l+1) - (l)}. \qquad [4.24]$$

indicating an identically valued reflectance function at $l$ and at $l+1$. The associated reflectance coordinates locate this "double extremum" which is in fact an artifact, converging toward a single point as the discrete wavelength step ($\epsilon$) decreases. Figure 4.6 shows a reflectance solid sectioned by a plane normal to the $x_1$ axis through its centroid. The facets have been colour randomly. The long edges common to

Figure 4.6: Reflectance Solid Edge (artifact)

two facets of similar orientation are sampling artifacts; with increased sampling the narrow, adjacent facets would continue to shrink.

### 4.2.10 Geometric Features of Reflectance Space

Features need not be artifacts of sampling. The black apex always exists and is a vertex at a non-differentiable point. It provides an initial point in exploring other surface features. Facets containing the origin necessarily retain one minima, which occurs at a wavelength governed by the common lower zone. The set of facets sharing the black apex may be arranged in a cycle (as with any set of polyhedron faces arranged about a common vertex). Adjacent faces share a common edge; the set of common edges forms a ray bundle emanating from the origin.

Motion along any edge away from the black apex increases the reflectance function except at those wavelengths fixed by the zones common to the edge (two, for the solid in three-space). The simultaneous minima need not be adjacent: the extrema present in the three basis vectors may allow them to lie at dissimilar wavelengths. In this case, a "crease" exists (a dihedral which does not approach 180° in the limit). Outward motion along this edge is along a linear path as only one free

parameter may be adjusted in the linear equation which define $\rho(\lambda)$; the other two are constrained to maintain the two extrema. A crease is evident along the upper right-hand edge of the solid appearing in Figure 4.6.

The convex hull of these edges form a "framework" for the conical end of the reflectance solid near regions of low reflectance. This end of the solid may be called the "dark cone". A conjugate "light" cone does not exist as there is no white apex. Instead, the reflectance function increases away from the origin until $m$ (three) maxima occur.

Where the facets of the dark cone intersect the rest of the solid's surface (the set of reflectance functions containing maxima) an edge set is created whose reflectance functions contain both minima and maxima. These reflectance functions are particularly important as they define saturated (Falk, Brill and Stork, 1986) colours which simultaneously include or exclude (as much as possible) regions along the spectrum having minimum and maximum reflectivity. These physical reflectances include the ideal *additive primary* colours red, green and blue (whose functions resemble low-pass, band-pass and high-pass filters) and more important, the *subtractive primary* or *secondary* colours cyan, magenta and yellow. The latter model idealized pigments (Feller, 1986) useful for printing and reproduct and have reflectances resembling band-stop functions. (Other colour systems defined in terms of their reflectances include the Munsell system (Munsell, 1963).)

These edges are formed along the intersection of planes lying at the lower and upper boundary of distinct zones. Since these adjacent planes are unrelated, their orientations are distinct. Once again, a crease of dimension $m-2$ is formed where the two hyperplanes of dissimilar orientation intersect. A point lying along a crease is indicated below, along with its companion reflectance function. A set of creases forms a partial path when arranged end-to-end. However, the lack of symmetry

prevents a closed *belt* – the edges may not be ordered cyclically. Instead, a blunt area is seen at the extreme right of Figure 4.3. (Belts are described again in §5.3.3 and at length in Chapter 7.)

Thee areas define regions of mixed minima and maxima for which motion within the zone allows nearby extrema to come into contact with the reflectance limits. While in contact, their defining zone is present, thereby introducing an additional facet to the solid. Thus, features related to the coordinates for which a lobe first makes contact with the physical limits forms dihedral breaks. Additional exploration reveals facets of appreciatable size associated with pronounced peaks in their associated reflectance functions. The edges of these areas contain multiple maxima. The location of these faces and their relation to other regions of the solid is not yet understood.

### 4.2.11 Advantages and Liabilities

The plane equations which define a reflectance solid generalize naturally to the $m$th dimension. The procedural determination of surface geometries (CAD/CAM in dimensions higher than three) do not. By representing a convex polytope in terms of its bounding hyperplanes and converting onto to a projective form when required, modeling software which allows for the cross-sectioning of four and higher-dimensional colour space may be created based upon first principles. Such software was created (see Appendix E) for the exploration of reflectance spaces described in Chapter 5.

The lack of both a symmetric white apex and a suitable definition for "approximate" white are liabilities, as are practical means of its computation. Irregularities of the reflectance solid, such as the "blunting" near the white end plus the seem-

ingly random location of dihedral creases are troublesome. These persist despite an increase in the sampling resolution of the solid.

The overall shape of the reflectance solid may be characterized in a small number of parameters by containment through a parallelotope. However, the presence of surface features not easily categorized limits the accuracy in establishing a tight fit.

## 4.3 Summary and Conclusions

A consistent model based on linear algebra has been extended to the realm of subtractive colour. The model and related notation directly embrace the additive system, allowing both to be treated within a common algebra. The use of orthogonal bases in the modeling of spectral reflectance functions is advocated. These span a Euclidean space, encouraging the study of the solids of reflectance space by geometric means. Unfortunately, these reveal the absence of a uniform white point, a distinct liability. A constraint which admits the white point while introducing geometric symmetry is treated in the following chapter.

# Chapter 5

# Symmetric Reflectance Space

In this chapter reflectance bases are suitably constrained to overcome their major liability: the inexact representation of the ideal white reflectance. The first section describes the modifications used to introduce this condition. The second further describes the symmetric solids so created in terms of algebra and analytic geometry. This section formally relates the modified reflectance solids to the *zonotopes.* (A brief review of the geometry $n$-dimensional Euclidean space and the zonotopes appears in Appendix B.) The chapter concludes with an exposition of prototypical reflectance solids and their related idealized reflectance functions.

## 5.1 Modified Reflectance Space

The set of orthogonal basis vectors $\beta_1(\lambda) \ldots \beta_m(\lambda)$ which span any reflectance space may be constrained such that the ideal white reflectance $\rho^{(\text{white})}(\lambda) = 1$ lies within the space. As will be seen, this introduces geometric symmetry while making little concession to overall representational accuracy.

Setting $\beta_1(\lambda) = c$ is sufficient to guarantee this condition. It is not necessary: the white reflectance may be found indirectly as a linear combination of a subset of the elements. In this case, an orthogonal *rotator* matrix (Watkins, 1991a) provides an alternate basis[1] manifesting the desired eigenvector $\beta_1(\lambda) = c$ while preserving orthogonality. Regardless, the modified reflectance spaces establish the element in the primary position of the basis set. This guarantees its presence in any reflectance space of arbitrary dimension $m$.

Notationally, $\bar{\beta}_1(\lambda) = 1$. The bar diacritical distinguishes the basis vector from its empirically-derived counterpart $\beta_1(\lambda)$ and suggests the former's *a priori* assignment to a constant reflectance. The remaining basis vectors are given the distinct notation $\tilde{\beta}_i(\lambda)$, with " ˜ " suggesting their approximating nature, as they are no longer the principle components of the data set, nor eigenvectors of the solid's covariance matrix.

Linear reflectance spaces were describe in the previous chapter. The defining equation for reflectance solids may now be modified:

$$\rho(\lambda) = x_1\bar{\beta}_1(\lambda) + x_2\tilde{\beta}_2 + \ldots + x_m\tilde{\beta}_m(\lambda) \qquad [5.1]$$

or

$$\rho(\lambda) = x_1 + x_2\tilde{\beta}_2 + \ldots + x_m\tilde{\beta}_m(\lambda) \qquad [5.2]$$

Substitution of the reflectance coordinates $\mathbf{x} = (c, 0, \ldots, 0)$ reveals the *grey reflectance* $\rho(\lambda) = c$, independent of choice of $\tilde{\beta}_2(\lambda)\ldots\tilde{\beta}_m(\lambda)$. The locus of points thus form a "grey" axis through the solid. That is, $\bar{\beta}_1(\lambda)$ is the body-diagonal $[1, \ldots, 1]$ of the parent $n$-hypercube lying in (unmodified) reflectance space.

[1]Traditionally, rotators are used to place zeros on a column, that is, force alignment to a coordinate axis (*ibid,* p. 140), not a diagonal.

### 5.1.1 Grey Axis Representation

Substitution of [5.2] into the physical constraint [4.2] gives

$$0 \leq x_1 + x_2\tilde{\beta}_2(\lambda) + \ldots + x_m\tilde{\beta}_m(\lambda) \leq 1. \qquad [5.3]$$

These limits are met exactly by suitable choice of reflectance coordinates. Clearly, $\mathbf{x} = (0, \ldots, 0)$ meets the lower limit for all $\lambda$. Likewise, setting $\mathbf{x} = (1, 0, \ldots, 0)$ meets the upper limit. In fact, on setting $x_1 = 1$, further substitution forces the remaining coordinates to zero. Since all are basis elements are orthogonal, the trailing set must also satisfy a defining requirement of independence:

$$x_2\tilde{\beta}_2(\lambda) + \ldots + x_m\tilde{\beta}_m(\lambda) = 0 \quad \Longleftrightarrow \quad x_2 = \cdots x_m = 0. \qquad [5.4]$$

This establishes the unique reflectance coordinates of the grey axis in any modified reflectance space.

### 5.1.2 Modified Orthogonal Bases

Orthogonality likewise implies that the projection of any non-initial basis element $\tilde{\beta}_i(\lambda)$ with $\bar{\beta}_1(\lambda)$ is zero. Thus, the areas (integral) under the curve defined by any $\tilde{\beta}_i(\lambda)$ is zero: they show no "DC bias". The basis functions thus exhibit both positive and negative regions, further suggesting the use of the "˜" notation and the distinction between $\tilde{\beta}(\lambda)$ and $\bar{\beta}(\lambda)$.

In the case of discrete vectors, the row sum of the elements of any $\tilde{\mathbf{b}}$ are zero save for $\bar{\mathbf{b}}_1$. These columns form $\mathbf{B}$, allowing the expression of the row summation using matrix multiplication on $\mathbf{B}^{\mathrm{T}}$:

$$(c, 0, \ldots 0)^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}(1, \ldots, 1)^{\mathrm{T}}, \qquad [5.5]$$

indicating that the grey axis projects merely onto the initial reflectance coordinate using the mapping defined in [4.5].

This identity suggests a means of creating a set of trailing basis elements $\tilde{\mathbf{b}}_i$ which are orthogonal to both each other and to initial element $\bar{\mathbf{b}}_1$. Construction of this set represents a point of departure from traditional methods of regression, such as principle component analysis. Here, one component is prescribed, while the remaining members are calculated empirically. This suggests a hybridized method which utilizes both *a priori* and *a posteriori* techniques.

### 5.1.3 Basis Construction

The construction of basis for modified reflectance space represent a point of departure from conventional regression methods. The introduction of an element useful to colorimetric modeling introduces a potential loss in the models ability to minimize variance for a given data set.

Two techniques are described which take advantage of conventional SVD to introduce the grey axis while enforcing orthogonality. The first method employs the Power method of SVD solution[2] to find dominant eigenvectors in succession, which are the principle components. Adjustments are introduced at each step to insure orthogonality. The second method preadjusts the input data to remove the presence of the grey axis, then proceeds along conventional lines.

The reorientation of the basis is tabulated for both methods, thus indicating loss and progressive recovery in minimization of variance as a function of dimension. Both approaches achieve the design objectives: only a slight increase in the dimension of the deflated subspace is required to produce a spectral approximation

[2]This is reviewed in Appendix A.

having high precision. However, liabilities in the first method leave the second the superior choice.

As before, the notation $\mathbf{S}$ describes a sample (data) matrix, $\mathbf{SS}^{\mathrm{T}}$ its covariance. The vector $\mathbf{v}$ is an eigenvector with associated eigenvalue $\mu$. More specifically, $\mathbf{b}$ describes an (unmodified) eigenvector – no diacritical is present. Similarly, $\bar{\mathbf{b}}_1$ is the constrained, constant vector, as suggested by the straight line; $\tilde{\mathbf{b}}$ is a unit vector orthogonal to $\bar{\mathbf{b}}_1$ (its elements sum to zero) with notation suggestive of its bipolar range. Also, $\mathbf{u}$ is a vector of all ones.

### 5.1.4 Successive QR Adjustment

The approach begins with the complete eigenvalue set $\mathbf{B}$ of the covariance matrix $\mathbf{C} = \mathbf{DD}^{\mathrm{T}}$ as calculated in §4.1.10. The first (basis) eigenvector $\mathbf{b}_1$ is replaced with the uniform, unit length grey basis vector $\bar{\mathbf{b}}_1 = (1, 1, \cdots, 1) \times (n)^{-\frac{1}{2}}$. The Gram-Schmidt process is then used to make the remaining basis vectors orthogonal. For instance, the second basis vector is created by projecting out the component parallel to the uniform grey vector, $\bar{\mathbf{b}}_1$ producing the intermediate $\mathbf{b}_2^*$:

$$\mathbf{b}_2^* = \mathbf{b}_2 - (\mathbf{b}_2 \cdot \bar{\mathbf{b}}_1)\mathbf{b}_2. \qquad [5.6]$$

Next, $\mathbf{b}_2^*$ is normalized to form the second basis vector $\tilde{\mathbf{b}}_2 = \mathbf{b}_2^* ||\mathbf{b}_2^*||^{-1}$. Successive basis vectors are likewise calculated by projecting out all previous vectors (including the first):

$$b_j^* = \mathbf{b}_j - \sum_{i<j} (\mathbf{b}_j \cdot \tilde{\mathbf{b}}_i)\mathbf{b}_j, \qquad [5.7]$$

and then normalizing:

$$\tilde{\mathbf{b}}_j = \mathbf{b}_j^* * ||\mathbf{b}_j^*||^{-1}. \qquad [5.8]$$

Note that while new basis $\mathbf{B} = [\bar{\mathbf{b}}_1|\tilde{\mathbf{b}}_2|\cdots|\tilde{\mathbf{b}}_n]$ resembles the original set $\mathbf{B} = [\mathbf{b}_1|\cdots|\mathbf{b}_n]$, it is *not* a set of eigenvectors and has no associated eigenvalues.

This process is analogous to $QR$ decomposition of a general matrix $\mathbf{M}$ in which $\mathbf{Q}$ is the resulting orthogonal basis, suggesting the MATLAB fragment:

```
%  input --  S a rectangular data matrix
%  output -- B the adjusted eigenvectors (columns)
   <r,c>   = size(S);
   <q,l,x> = SVD(S);
   v(:,1)  = ones(r,1)/sqrt(r);
   <B,x>   = QR(v);
```

In essence, the original basis — which is derived using conventional eigenvalue methods — is rotated through small angles so that its first eigenvector is coincident with the desired ones vector $\bar{\mathbf{b}}_1$. The approach is accurate only inasmuch as the rotation is small, otherwise the necessary rotation destroys goodness of fit. The method succeeds because in practice $\beta_1(\lambda) \approx \bar{\beta}_1(\lambda) = 1$ often holds.

Goodness of fit may be examined by studying the projected length of the new basis matrix $\tilde{\mathbf{B}}$ element against the original eigenvector basis $\mathbf{B}$. The Cartesian product of these columns is $\mathbf{B}^{\mathrm{T}}\tilde{\mathbf{B}}$, yielding a matrix whose element $(i, j)$ is the direction cosine of the angle formed between the original element $\mathbf{b}_i$ and modified element $\tilde{\mathbf{b}}_j$. That is, a study of *principal angles* of *principal vectors* is conducted (Watkins, 1991b). When the approximate condition $\mathbf{b}_i \approx \tilde{\mathbf{b}}_j$ is met exactly, unit elements appear along the diagonal. Similarly, the presence of small off-axis elements indicates a $n$-space rotation in which only small perturbations occur. The data derived from the Kodak SVC6500 printer appear in Table 5.1. This method is easily accommodated by sequential SVD algorithms. Conceptually, the supplanting

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0.981 | 0.192 | 0.016 | -0.035 | 0. | -0.008 | 0.006 | -0.002 |
| 2 | -0.187 | 0.977 | -0.052 | 0.088 | 0. | 0.019 | -0.014 | 0.004 |
| 3 | -0.023 | 0.045 | 0.998 | 0.043 | 0. | 0.008 | -0.006 | 0.002 |
| 4 | 0.049 | -0.077 | -0.037 | 0.982 | 0.006 | -0.125 | 0.092 | -0.024 |
| 5 | -0.001 | 0.001 | 0. | -0.005 | 1. | 0.003 | -0.002 | 0. |
| 6 | 0.005 | -0.008 | -0.004 | 0.038 | -0.001 | 0.778 | 0.624 | -0.055 |
| 7 | -0.006 | 0.008 | 0.004 | -0.038 | 0.001 | -0.220 | 0.358 | 0.906 |
| 8 | -0.021 | 0.031 | 0.013 | -0.139 | 0.002 | -0.569 | 0.685 | -0.418 |

Table 5.1: Basis Rotation: SVD versus QR

action of $\bar{\mathbf{b}}_1$ upon $\mathbf{b}_1$ is the introduction of bias upon the latter. When this grows large, the projected length of $\bar{\mathbf{b}}_1$ onto $\mathbf{b}_1$ becomes small and the angle between coordinate systems causes them to diverge. For small vectors, this residual may be great enough to greatly alter the orientation of the original and treated basis vector. This is seen at table entry $[8, 8]$ above, where the negative component $-.418$ indicates that the angle is $114.7°$.

A fully general approach would employ the power method implicitly in finding the dominant spectral components of the unmodified covariance matrix $\mathbf{C}$ in iterative fashion, with orthogonality enforced at each step as described above. This operation is hampered in finding a suitable "pseudo"-eigenvalue $\mu$ associated with the grey axis "pseudo"-eigenvector $\mathbf{v}$. This pair is essential in performing the residual deflation step on the covariance $\mathbf{S}' = \mathbf{S} - \mu\mathbf{v}\mathbf{v}^{\mathrm{T}}$ and is additionally sensitive to choice of $\mu$. In particular, $\mathbf{S}'$ might take on negative eigenvalues. The "constrained eigenvector" problem arising from the research presented here is seldom encountered elsewhere and is considered by experts in the field to be an interesting topic

in its own right (Overton, 1990).

### 5.1.5 Preadjusting Modification (MSSM)

The drawbacks of the previous method suggest a preadjustment to the data to remove the constraint. In this second method, removal of the grey component *precedes* the SVD algorithm. Here the column sums of the data matrix $\mathbf{S}$ are forced to zero by subtracting the mean value of each column from that column's elements. (This offsets the mean of each trailing column to zero while preserving the variance and higher order statistical *moments* of each row). The formula which maps column vector $\mathbf{s}_i$ onto $\mathbf{s}'_i$ resembles Gram-Schmidt:

$$\mathbf{s}'_i \leftarrow \mathbf{s}_i - (\mathbf{s}_i \cdot \mathbf{u})\mathbf{u} \qquad [5.9]$$

where $\mathbf{u} = (1, 1, \cdots, 1) \times n^{-1}$ which serves as a uniform grey reflectance sample whose components sum to one (hence $\bar{\beta}_1(\lambda) = 1$ for all $\lambda$, but $\bar{b}_1[l] = n^{-\frac{1}{2}}$, thus having unit length but not unit elements). The resulting sample matrix $\mathbf{S}'$ has been preadjusted so that one degree of freedom (the "DC offset") has been removed. This reduction implies that the related covariance matrix is reduced by one in rank. In particular, it is now singular and thus has no inverse. This presents no difficulty to the operation of the SVD.

The summing action of [5.9] across on all columns may be conveniently represented using the machinery of matrix multiplication, as in [5.5]. The tallying of column sums into a row vector is created by premultiplying the unmodified samples $\mathbf{S}$ with $\mathbf{u}$'s transpose, that is, $\mathbf{u}^{\mathrm{T}}\mathbf{S}$. Duplicating rows to achieve conformality with $\mathbf{S}$ requires premultiplication by $\mathbf{u}$, leaving the adjustment value $\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{D}$, so that subtraction gives

$$\mathbf{S}' = \mathbf{S} - \mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{S} = (\mathbf{I} - \mathbf{u}\mathbf{u}^{\mathrm{T}})\mathbf{S} = \mathbf{M}\mathbf{S}, \qquad [5.10]$$

in which $\mathbf{M} = (\mathbf{I} - \mathbf{u}\mathbf{u}^{\mathrm{T}})$ represents the matrix of modification. (This class of matrix closely resembles the Householder transformation (Wilkinson, 1965).) The derivation of this general form may be confirmed by verifying the removal of the equipotential component $\mathbf{u}$:

$$\begin{aligned} \mathbf{u}^{\mathrm{T}}(\mathbf{M}\mathbf{S})\mathbf{u}^{\mathrm{T}}\mathbf{S}' &= \mathbf{u}^{\mathrm{T}}\mathbf{S} - \mathbf{u}^{\mathrm{T}}\mathbf{u}\mathbf{u}^{\mathrm{T}}\mathbf{S} \\ &= \mathbf{u}^{\mathrm{T}}\mathbf{S} - (1)\mathbf{u}^{\mathrm{T}}\mathbf{S} \\ &= (0)\mathbf{S} \\ &= \mathbf{0}, \end{aligned} \qquad [5.11]$$

where $\mathbf{u}^{\mathrm{T}}\mathbf{u} = 1$, that is, $\mathbf{u}$ is an arbitrary unit vector. Using this approach, the remaining principle components are the eigenvectors of the preadjusted covariance $\mathbf{C}' = \mathbf{S}'\mathbf{S}'^{\mathrm{T}}$. Because $(\mathbf{X}\mathbf{Y})^{\mathrm{T}} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}$ for any $\mathbf{X}$ and $\mathbf{Y}$ and because $\mathbf{M} = (\mathbf{I} - \mathbf{u}\mathbf{u}^{\mathrm{T}})$ is symmetric, the eigenpair problem on $\mathbf{C}'$ may be modeled as:

$$\begin{aligned} \mathbf{C}' &= \mathbf{S}'\mathbf{S}'^{\mathrm{T}} \\ &= (\mathbf{M}\mathbf{S})(\mathbf{M}\mathbf{S})^{\mathrm{T}} \\ &= \mathbf{M}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{M} \\ &= \mathbf{M}\mathbf{S}\mathbf{S}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}} \\ &= \mathbf{M}\mathbf{C}\mathbf{M}^{\mathrm{T}} \\ &= \mathbf{M}\mathbf{C}\mathbf{M}, \end{aligned} \qquad [5.12]$$

closely resembling the original problem on $\mathbf{C} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$. The penultimate term in [5.12] indicate that $\mathbf{C}$ undergoes a *similarity transformation.* In fact, $\mathbf{M}$ is an *idempotent* matrix[3] so that $\mathbf{M}^n = \mathbf{M}$; this is clear because repeated rank-reductions on arbitrary $\mathbf{P}$ cannot remove a projected component (here the grey basis vector) once it is eliminated. That is,

$$\begin{aligned} \mathbf{M}^n\mathbf{P}\mathbf{M}^n &= \mathbf{M}^{n-1}(\mathbf{M}\mathbf{P}\mathbf{M})\mathbf{M}^{n-1} \\ &= \mathbf{M}^{n-1}\mathbf{P}\mathbf{M}^{n-1} \\ &= \mathbf{M}\mathbf{P}\mathbf{M}, \end{aligned} \qquad [5.13]$$

[3]This is also called a *orthoprojector* matrix and resembles the form appearing in [3.22].

with the last equality formed through induction. The adjustment $\mathbf{MSS}^{\mathrm{T}}\mathbf{M}$ to data matrix $\mathbf{S}$ suggests the name appearing in the section title.

This form closely resembles the eigenpair preserving *similarity transformation* $\Lambda\mathbf{A}\Lambda^{\mathrm{T}}$ commonly employed in solving eigensystems (Acton, 1970) except here $\mathbf{M}$ is not orthogonal, so that $\mathbf{M}^{\mathrm{T}} = \mathbf{M}^{-1}$ cannot hold. Instead, the form has a covariance having $n-1$ unit eigenvalues and is often called a *reflector* matrix (Watkins, 1991a).

Implementation of this reduction were also conducted and are outlined in the MATLAB fragment, below. Note in particular the parenthesized subexpression `(u'*S)` appearing on the third line.

```
<r,c> = size(S);
u = ones(r,1);
S2 = S - (1/r)*u*(u'*S);
<B,w,x> = svd(S2);
tmp = B(:,1);
for i=1:r
  tmp2 = B(:,i), B(:,i) = tmp, tmp = tmp2;
  end;
B(:,r) = tmp;
```

Since matrix produces may be reassociated freely, this order of evaluation is employed to achieve a speed-up on the largest data sets encountered by minimizing the size of intermediate matrix products (Santoro, 1984). Assuming the product $\{\mathbf{m}_{ij}\} = \{\mathbf{a}_{ik}\}\cdot\{\mathbf{b}_{kj}\}$ has a cost proportional to $ijk$, then the cost of left-associative evaluation of a $[512 \times 4096]$ data set $\mathbf{S}$ (the largest to date, see Chapter 8) is $(2^{18}+1)(2^{12})$. This is reduced to $(2)(2^{9})(2^{12})$, a ratio slightly more than $2^{8} = 256$ when regrouped.

The permutation loop which concludes the code requires an explanation. Applying the SVD to `S2` necessarily yields a zero eigenvalue in its final position, corre-

sponding to the singularity introduced through loss of rank by $\mathbf{M}$. More important, the (grey) basis vector $\mathbf{u}$ reappears in association with this eigenvalue, as this axis lies fully within the null space of $\mathbf{S}'$. (Zero eigenvalues still have an associated unit eigenvector.) Circulation of the matrix to the right returns this trailing column to first position; the remaining vectors are associated with eigenvalues of decreasing value. Unlike the previous method, the values associated with each vector are true eigenvalues, but in reference to a modified covariance matrix $\mathbf{MCM}$.

For partial SVD decomposition of large data sets – as previously advocated – the final $\mathbf{u}$ remains uncalculated, requiring the grey basis vector to augment the eigenvector on the left-hand side. Alternately, the partial SVD may run to conclusion and the effects of numerical roundoff studied by comparing the final eigenvector against $\mathbf{u}$, a useful property. (See §6.4.4 and in particular, the third column Table 6.1.)

The SVC6500 printer data was calculated using this method with resulting basis projected onto the unadjusted eigenvector basis, as before. The matrix of direction cosines appears in Table 5.2. Its elements are comparable but dissimilar to those of the previous method.

A diagonal element of $-1$ indicates a reflection has taken place in reference to the associated element of the unmodified colour space, since $\cos(-1) = 180°$. The vectors are thus parallel, the handedness of the coordinate system reversed without altering the geometry of the reflectance solid in the modified space. The lower $6 \times 6$ block is very nearly the identity, indicating that the "slack" lost in eliminating an axis which was not a principle component (a true eigenvector) quickly taken up.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | -0.981 | -0.187 | -0.024 | -0.05 | -0.001 | -0.006 | 0.016 | 0.021 |
| 2 | 0.187 | -0.982 | 0. | 0. | 0. | 0. | 0. | 0. |
| 3 | 0.023 | 0.004 | 1. | 0. | 0. | 0. | 0. | 0. |
| 4 | -0.049 | -0.009 | -0.001 | 0.999 | 0. | 0. | 0. | 0. |
| 5 | 0.001 | 0. | 0. | 0. | -1. | 0. | 0. | 0. |
| 6 | -0.005 | -0.001 | 0. | 0. | 0. | 1. | 0. | 0. |
| 7 | 0.006 | 0.001 | 0. | 0. | 0. | 0. | 1. | 0. |
| 8 | 0.021 | 0.004 | 0.001 | 0.001 | 0. | 0. | 0. | 1. |

Table 5.2: Basis Rotation: SVD versus MSSM

### 5.1.6 Comparisons

A simple comparison of SVD, QR, and MSSM methods plots the chromaticities of the elements produced by each produced by each. A plot using the LUV space of colour differences appears in Figure 5.1. (That is, the perceptual chromaticities of each model reflectance is considered under illumination by Source E).

As seen in the figure, the first element of both the QR and MSSM methods coincide at the chromaticity of source E (see also Figure 3.3), in proximity to the nearly grey reflectance $\tilde{\beta}_1(\lambda)$ derived through SVD. This introduces a deviation in the chromaticities of the second basis elements, which is nearly gone by the third.

Although the method of QR adjustment provided slightly superior goodness of fit for a few trial data sets, the presence of negative eigenvalues among the trailing basis elements are its downfall. These reveal the impossibility of locating a dominant component in a data set while simultaneously forcing it into orthogonality with the previous set of basis vectors. By comparison, the MSSM method presents

Figure 5.1: Basis Element Chromaticities (SVD, QR, MSSM)

no computational difficulties: the symmetric forms encountered are all semi-definite (the eigenvalues are all non-negative). This allows for a full decomposition with positive eigenvalues, though neither the related eigenvectors are principle components, nor do the associated eigenvalues directly gauge goodness of fit, though for small changes the rotation matrix tabulated in Table 5.2 indicate small turning angles. The MSSM method is used exclusively for all work which follows.

## 5.2 Symmetric Reflectance Solids and Zonotopes

The modified bases in colour reflectance space gives rise to a new class of solids. As is demonstrated below, this change is both necessary and sufficient to change the nature of the related colour solids into generalized *zonotopes.* When modified, the reflectance solid $\mathcal{R}^n$ resembles the zonotope $\mathcal{Z}^n$. Zonotopes are of central importance to the thesis. They enjoy a strong strong history dating to the exploration

of symmetry classes by crystallographers in the late 19th century (Fedorov, 1971), and have a strong following among modern-day computational geometers (Coxeter, 1961), (Coxeter, 1962), (Coxeter, 1968), (Coxeter, 1973), (Heckbert, 1985). Their high degree of symmetry simplifies theoretical and practical models concerning them.

Most important, zonotopes provide a consistent geometrical framework for a series of reflectance solids in any dimension. The presence of the component $\bar{\beta}_1(\lambda)$ in any modified reflectance space guarantees an axis through the centroid of the parent $n$-hypercube which ultimately defines the space. Because the central section of a zonotope remains a zonotope (Coxeter, 1973), any lower dimensional model retains this attribute. In addition, the (location of) the centroid is preserved, allowing exact representation of the grey point and associated axis in solids of deflated dimension.

This suggests an ordered sequence of reflectance solids which define a colour space at increasing precision. Each left-hand set is a strict subset of sets appearing to the right:

$$\mathcal{R}^1 \subset \mathcal{R}^2 \subset \ldots \subset \mathcal{R}^n \qquad [5.14]$$

in which the one-dimensional reflectance solid $\mathcal{R}^1$ is the zonotope $\mathcal{Z}^1$ of unit interval defining overall reflectance (what an astronomer would call *albedo)* and the final reflectance solid $\mathcal{R}^n$ the zonotope hypercube $\mathcal{Z}^n$, providing a discrete representation of any valid reflectance function.

The (hyper) cube is the prototypical zonotope; all modified reflectance solids can thus be constructed as central sections of this parent reflectance solid in $\mathcal{R}^n = \mathcal{Z}^n$. Considering [5.14] in the other (left to right) direction, any low-dimensional reflectance space may be *inflated* onto a higher dimensional space of representation

by preserving these organizing principles. This is of practical value, as when the reflectance coordinates are imaged on device having a larger number of printing channels. (Both inflation and this specific example are discussed at greater length in Chapter 7.) In short, *when inflated to their limiting dimension, symmetric reflectance solids become hypercubes, which are prototypical zonotopes.*

The modification is effected by constraining the initial basis element $\beta_1(\lambda)$ to a constant, forming $\bar{\beta}_1(\lambda)$. Substituting $x_1 = x_1' + \frac{1}{2}$ into the constants already present in [5.3] displaces the solid relative to its original reference frame while leaving its underlying geometric features unaltered. The spatial displacement imparts symmetry to the algebra of the plane equations:

$$\frac{1}{2} \leq x_1' + x_2\tilde{\beta}_2(\lambda) + \ldots + x_m\tilde{\beta}_m(\lambda) \leq \frac{1}{2}, \qquad [5.15]$$

which may be recast after [4.17] as an intersection of zones containing the origin:

$$\mathcal{R}^n \equiv \bigcap \left[ \|x_1' + \sum_{i=2}^{n} x_i\beta_i(\lambda)\| \leq \frac{1}{2} \right], \qquad [5.16]$$

in which the latter suggests the solid's central symmetry, described shortly. In terms of colour, the limiting points for black and white are now present in symmetric arrangement on any modified solid. These have coordinates $(0, 0, \ldots, 0)$ and $(1, 0, \ldots, 0)$, defining the endpoints of the grey axis. Generally, the solid common to the set of zone intersection cannot be empty, as it least this axis will be present, as in $Z^1$.

### 5.2.1 Reflectance Solids as Zonotopes

The algebra and geometry of the previous two sections may be brought into agreement by straightforward application of analytic geometry. The proof outlines offered below are specific to the colour solids defined in this thesis. Companion entries

of a more general nature appear in Appendix B. For formal proofs, the reader is referred to standard treatments (Coxeter, 1973), (Grünbaum, 1967).

**Theorem 1.** *Modified Reflectance Solids Describe N-dimensional Zonotopes*

We assert that the convex polytope $\mathcal{R}^n$ defined by the locus of points in $n$-space $x = (x_1, \cdots, x_n)$ which satisfy

$$0 \leq \sum x_i \beta_i(\lambda) \leq 1 \qquad [5.17]$$

is centrally symmetric *iff* the vector $\beta(\lambda) = 1$ lies within the space.

**Proof.** A convex polytope is a zonotope *iff* all points $x$ on its surface are centrally symmetric about the solid's centroid $x^{(\mathrm{c})}$ (Coxeter, 1973). Here $x^{(\mathrm{p})} = x^{(\mathrm{c})} + x^{(\mathrm{d})}$ describes the point on the surface with $x^{(\mathrm{d})}$ the displacement from the centroid. The point lies on the limit of a zone(s). Then $x^{(\mathrm{p}')} = x^{(\mathrm{c})} - x^{(\mathrm{d})}$ necessarily describes an paired, antipodal point also on the surface on the same zone's opposite boundary. The name *conjugate* is chosen to describe such points. The plane equation for the zone facet through point $x$ is:

$$(x_1^{(\mathrm{c})} + x_1^{(\mathrm{d})})\beta_1(\lambda) + (x_2^{(\mathrm{c})} + x_2^{(\mathrm{d})})\beta_2(\lambda) + \cdots + (x_n^{(\mathrm{c})} + x_n^{(\mathrm{d})})\beta_n(\lambda) = 0. \qquad [5.18]$$

Likewise, the companion plane forming the zone at conjugate $x'$ is described by:

$$(x_1^{(\mathrm{c})} - x_1^{(\mathrm{d})})\beta_1(\lambda) + (x_2^{(\mathrm{c})} - x_2^{(\mathrm{d})})\beta_2(\lambda) + \cdots + (x_n^{(\mathrm{c})} - x_n^{(\mathrm{d})})\beta_n(\lambda) = 1. \qquad [5.19]$$

Choosing $x^{(\mathrm{c})}$ to satisfy the former must allow the latter to hold for arbitrary $x^{(\mathrm{d})}$. Adding both, it is then necessary that

$$2x_1^{(\mathrm{c})}\beta_1(\lambda) + 2x_2^{(\mathrm{c})}\beta_2(\lambda) + \cdots + 2x_n^{(\mathrm{c})}\beta_n(\lambda) = 1 \qquad [5.20]$$

hold universally. Because $\beta_1$ is a constant the $\lambda$ argument may be dropped and $\beta_1'$ used to indicate the value of this constant. Finally, setting $c_1 \ldots c_n = 0$ removes

those terms which vary as a function of wavelength, leaving

$$2c_1\beta_1' = 1 \qquad [5.21]$$

or

$$x^{(c)} = (\frac{1}{2\beta_1'}, 0, \ldots, 0), \qquad [5.22]$$

thus establishing the reflectance coordinates of the centroid. In discrete form, the coordinates are $x^{(c)} = (\frac{1}{2}n^{\frac{1}{2}}, 0, \cdots, 0)$, with the $n$ appearing because the first reflectance vector $\bar{\mathbf{b}}_1(l) = [1, \cdots, 1] \times n^{-\frac{1}{2}}$ is scaled (see also [5.44]) in order to preserve unit length. ∎

**Corollary 2.** *No Other Modification to Reflectance Solids Yields Zonotopes.*

**Proof.** This falls out immediately from Theorem 1. Exactly one term can be permitted in the left-hand side of [5.17] to allow the equation to hold under arbitrary $\beta$ and $x$, from which Theorem 1 is based. Any additional $\beta$ elements similarly constrained are linearly dependent, in violation of basis orthogonality. Thus, constraining the first basis element is necessary and sufficient to assure that reflectance solids are zonotopes in reflectance spaces of arbitrary dimension. ∎

Central symmetry provides the key distinction between modified and unmodified reflectance solids. The single modification is significant: further *a priori* constraints on the basis elements are contraindicated, as symmetry would then be destroyed. Adjustments made possible by relaxing the conditions of basis orthogonality are discussed in §8.3.2.

It is thus sufficient though not necessary that $\beta_1(\lambda)$ contain constant entries to form a space in which reflectance solids are zonotopes, as the grey axis may be formed through linear combination of multiple basis elements. In the latter case the solid is merely rotated, as described at the beginning of the chapter. Since

the one-dimensional modified reflectance space must contain the grey reflectance $\rho(\lambda) = 1$, the equipotential element $\bar{\beta}_1(\lambda)$ must be installed in first position in order for the nesting in [5.14] to hold. Therefore, the family of modified reflectance solids are the zonotopes $\mathcal{Z}^1 \cdots \mathcal{Z}^n$ *iff* some reflectance $\rho(\lambda) = c$ lies within the space.

### 5.2.2 Conjugates and Complementary Colours

All (in)valid points $x^{(\mathrm{p})}$ in modified reflectance space have a *conjugate* (in)valid point $x^{(\mathrm{p}')} = x^{(\mathrm{c})} - x^{(\mathrm{p})}$. Algebraically, any reflectance $\rho(\lambda)$ has a conjugate, complementary reflectance $\rho'(\lambda)$ found by

$$\rho'(\lambda) = 1 - \rho(\lambda). \qquad [5.23]$$

From this it is easily seen that the conjugates are reflexive

$$\rho''(\lambda) = 1 - [1 - \rho(\lambda)] = \rho(\lambda). \qquad [5.24]$$

Since the centroid is its own conjugate,

$$\rho^{(\mathrm{c}')} = \rho^{(\mathrm{c})}(\lambda), \qquad [5.25]$$

so substitution of [5.25] into [5.23] and clearing sides gives

$$\rho^{(\mathrm{c})} + \rho^{(\mathrm{c})} = 1 \qquad [5.26]$$

or

$$\rho^{(\mathrm{c})} = \frac{1}{2}, \qquad [5.27]$$

as expected.

The centroid describes the (mid-level) *grey reflectance.* All other points occur in conjugate pairs symmetric about this *grey point.* The conjugate points describe reflectance functions of complementary colour.

The surface geometry of the reflectance solid is described uniquely by its convex hull. This is a set of $2h$ conjugate vertices related spatially through *central inversion.* The solid's surface and associated limiting functions may be probed interactively using the custom solid modeler described in Chapter 7. The tool readily depicts the mirroring of any reflectance function about the horizontal line $y = \frac{1}{2}$ when the point of selection in modified reflectance space moves to the conjugate point.

The term *opponent reflectance* function is introduced to describe the reflectance function of complementary colour associated with the conjugate points. This is by extension of the term *opponent colour,* which describes the appearance of the spectral power distributions which when combined form a white source (in some subspace of metamerism, as when their chromaticies coincide with source E when diagrammed). Opponent reflectance functions satisfy a stronger condition. They sum to form the function $\rho^{(\text{white})}(\lambda)$, a restatement of [5.23] achievable in any modified reflectance space. Thus, opponent reflectances are dissimilar even on the parent $n$-space, hence their reflectance coordinates are distinct in any reflectance space. The only exception is the self-conjugate grey centroid, including all higher-dimensional "grey" points sharing the reflectance coordinates $(\frac{1}{2}, 0, \ldots, 0)$. (That is, sources of black light[4] may be superimposed on the self-conjugate grey.) Opponent reflectances are encountered in an empirical situation in §7.3.3 and again in §8.3.2, where a plot is also presented (Figure 8.11).

### 5.2.3 Zonotopes as Reflectance Solids

**Lemma 3.** *A Zonotope Does Not Always Define a Reflectance Solid.*

[4]See §3.1.7 and especially §6.1.1 subsection "Null Space and Black Reflectance".

That is, the converse of Theorem 1 does not hold, Corollary 2 notwithstanding. An arbitrary $n$-zonotope, centered about the origin may be modeled by a set of plane equations:

$$\mathcal{Z}^n = \{\forall\lambda : \sum_{i=1}^{n} x_i\beta_i(\lambda) = \pm c_i\} \qquad [5.28]$$

having pairs of $(n-1)$-cell "faces" for each discrete $\lambda$. When a modified reflectance solid is translated to the origin it has the form

$$\forall\lambda : \sum_{i=1}^{n} x_i\beta_i(\lambda) = \pm\tfrac{1}{2}, \qquad [5.29]$$

which indicates that zonotopes are more general. ∎

The previous analysis suggests that all zones within a reflectance solid have a similar span (thickness), provided that adjustments to the constant appearing on the right-hand side of [5.28] (owing to plane equation normalization) are not large. When all thicknesses are similar, a hypersphere is suggested. This intuitive result is made rigorous below[5].

**Theorem 4.** *All Reflectance Solid are Contained Within a Circumscribing Hypersphere.*

A sphere of sufficient size naturally contains a finite collection of solids. However, a unique sphere exists which exactly contains (is tangent at two or more points for each element of) the entire set of reflectance solids. Recalling equation [4.16]:

$$0 \le x_1\beta_1(\lambda) + x_2\beta_2(\lambda) + \cdots + x_n\beta_n(\lambda) \le 1, \qquad [5.30]$$

then squaring and grouping cross-terms yields:

$$0 \le x_1^2\beta_1^2(\lambda) + x_2^2\beta_2(\lambda) + \cdots + x_n^2\beta_n^2(\lambda) + 2[x_1\beta_1(\lambda) + \cdots + x_n\beta_n(\lambda)] \le 1. \qquad [5.31]$$

[5]This is a necessary step: objects of uniform breadth need not be spherical. The most extreme counterexample is the regular tetrahedron whose four faces are replaced with four spherical caps whose centres lie at the vertex opposing each cap.

Integration of both sides with respect to $d\lambda$ then yields:

$$0 \leq x_1^2 \int \beta_1^2(\lambda) + \cdots + x_n^2 \int \beta_n^2(\lambda) + 2 \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} x_i x_j \int \beta_i(\lambda)\beta_j(\lambda) \leq c, \qquad [5.32]$$

in which $c$ is the constant of integration. For definite integrals across the desired spectral range this is $\lambda_{end} - \lambda_{start}$ and corresponds to the area of the unit rectangle in this range. By orthogonality of the set $\{\beta_1 \ldots \beta_n\}$ we may invoke the Kronecker delta,

$$\forall \lambda : \int \beta_i(\lambda)\beta_j(\lambda)\ d\lambda = \delta(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \qquad [5.33]$$

causing all cross-terms to vanish:

$$0 \leq x_1^2 \int \beta_1^2(\lambda) + \cdots + x_n^2 \int \beta_n^2(\lambda) \leq c \qquad [5.34]$$

which is

$$0 \leq x_1^2 c + \cdots + x_n^2 c \leq c. \qquad [5.35]$$

Finally, $c$ must be positive, else [5.35] describes the empty set. Clearing will not reverse the order of the inequality, yielding the equation for a *unit hypersphere,*

$$x_1^2 + \cdots + x_n^2 \leq 1, \qquad [5.36]$$

when the equality is met. This result makes additional containment heuristics §4.2.4 available to modified reflectance solids.

**Lemma 5.** *The Reflectance Solid is the Hypercube when $\beta(\lambda)$ are Diagonal Functions.*

In this case the Kronecker delta condition for orthogonality [5.33] holds for any fixed parameter $\lambda_p$, for at this wavelength the basis element $\beta_p(\lambda)$ will be unity, all others zero.

Intuitively, this forms a set of cutting zones of unit thickness with no normalization step required. Proceeding from [5.32] with this additional information yields:

$$0 \leq x_1^2\beta_1^2(\lambda) + \cdots + x_n^2\beta_n^2(\lambda) \leq C \qquad [5.37]$$

in which cross terms again drop out. The wavelength argument is thus fully preserved. Dropping any term $\beta_p(\lambda)$ when index $i \neq \lambda_p$ and choosing successive values of $\lambda$ thus yields a set of $2n$ inequalities:

$$0 \leq x_i^2\beta_i^2(\lambda)\ c \leq c \qquad [5.38]$$

for each value $\lambda_1 \ldots \lambda_n$. Because $\beta_i(\lambda) \geq 0$ for all $i$ and $\lambda$, all $x$ weighting coefficients must be positive to assure that $0 \leq \beta(\lambda) \leq 1$. This makes the inequality $0 \leq$ on the left-hand side of [5.38] redundant, leaving the final form

$$x_i \leq +1. \qquad [5.39]$$

Here $x_p$ assumes the value $+1$ at $\lambda_p$ of basis element $\beta_p(\lambda)$, giving it a unit length. The remaining coordinates must then be zero, else containment (Theorem 4) would be violated. Vertex elements of the form $(1, 0, \ldots, 0)$ permuted cyclically define the face normals of the unit $n$-hypercube of with $2^n$ vertices $(\frac{1}{2} \pm \frac{1}{2} \cdots \frac{1}{2} \pm \frac{1}{2})$. As expected, the reflectance cube has a centroid at $x^{(c)} = (\frac{1}{2} \cdots \frac{1}{2})$. ∎

Thus, both hypercubes and hyperspheres are zonotopes having geometric shape related to their metrical properties. (These are the Max or Euclidean norms, $\| \ \|^\infty$ and $\| \ \|^2$, respectively.) Either may be used to provide containment heuristics.

## 5.3 Prototypical Symmetric Reflectance Solids

A set of prototypical solids may be constructed based on sets of orthogonal bases **B** whose elements are prescribed *a priori.* Four solids are described in turn.

### 5.3.1 Colour Cubes

The most general choice of $\mathbf{B}$ must satisfy

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{I} \qquad [5.40]$$

to allow both [4.3] and [5.5] to hold exactly within any reflectance space. (This allows reflectance functions and reflectance coordinates to be mapped losslessly in either direction.) Then $\mathbf{B}^{\mathrm{T}} = \mathbf{B}^{-1}$, a defining condition for $\mathbf{B}$ being square and orthonormal. Thus, any alternate choice of $\mathbf{B}$ defines a rigid rotation or mirroring; no scaling or shearing is present. (As shown above, the solid is unique – only its form need be determined.)

The canonical colour solid of parent $n$-space may be found by again choosing the $n \times n$ identity as the basis $\mathbf{B}$ on this space. Substitution into the reflectance law [3.2] (taken in discrete form), yields

$$\forall i: \quad 0 \leq \mathbf{x}[i] \leq 0, \qquad [5.41]$$

valid for any $n$-space. That is, the unique[6] $n$-reflectance solid of valid physical colours is the *hypercube.* It has $2n$ faces determined by the $2^n$ vertices which are the extreme points determining its convex hull. The coordinates of the vertex are easily enumerated by counting in binary fashion. For instance, the vertices of the hypercube in three-space has coordinates

$$(0\ 0\ 0),\ \ (0\ 0\ 1),\ \ (0\ 1\ 0),\ \ (0\ 1\ 1),\ \ (1\ 0\ 0),\ \ (1\ 0\ 1),\ \ (1\ 1\ 0),\ \ (1\ 1\ 1). \qquad [5.42]$$

With $n = 3$, the associated reflectance tines may be considered non-overlapping bandpass reflectances spanning the red, green and blue spectral regions of primary

[6]Right and left-handed enantiomorphs are considered identical.

colours. The three antipodal vertices then define the complementary secondary colours cyan, magenta and yellow. Finally, the white (black) reflectance completes the set. It may be considered the sum of all (none) of the reflectances defining the space's axes. Regardless of the definition of the range of colour of the primaries, these last have well-defined reflectance functions $\rho(\lambda) = 1$ and $\rho(\lambda) = 0$, respectively. The points define the *body diagonal* of the colour cube. The segment is the locus of points whose related reflectance functions are the *idealized grey reflectance* $\rho(\lambda) = c$ with $0 \leq c \leq 1$. The corresponding coordinates are (c, ..., c). An illustration appears later in Figure 5.2(a) and (b).

### 5.3.2 Modified Cube

Although the unmodified cube is a zonotope, it does not contain the grey axis as its first basis element, in violation of [5.14]. The presence of the grey axis may be guaranteed by providing it in the first basis element in a space of three elements:

$$\bar{\beta}_1(\lambda) = 1. \qquad [5.43]$$

In the discrete setting,

$$\bar{\mathbf{b}} = [1, 1, \cdots, 1] \times (n)^{-\frac{1}{2}} \qquad [5.44]$$

such that the orthonormality $\bar{\mathbf{b}} \cdot \bar{\mathbf{b}} = 1$ holds.

The two remaining elements $\beta_2(\lambda)$ and $\beta_3(\lambda)$ must satisfy the Kronecker delta to provide orthogonality. In particular

$$\bar{\beta}_1(\lambda) \cdot \beta_2(\lambda) = \bar{\beta}_1(\lambda) \cdot \beta_3(\lambda) = 0 \qquad [5.45]$$

so that in discrete form the row sums of $\mathbf{b}_2$ and $\mathbf{b}_3$ are zero. As a further consequence, some elements of both will be negative, as will any other basis vector.

Figure 5.2: Colour Cubes

This forms a point of departure between reflectances $\rho(\lambda)$ and reflectance elements $\beta(\lambda)$: under the $\bar{\beta}_1(\lambda)$ condition the remaining basis elements necessarily describe non-physical reflectance. (An analogous condition is the lack of intersection between the additive colour solid and the axes of tristimulus space in Figure 3.2.) Geometrically, the modified colour solid may be depicted as a cube (Figure 5.2(a)) in which the initial axis is now the body diagonal. Equation [5.45] guarantees that the additional elements are orthogonal to this element and thus lie within the planar cross-section appearing in Figure 5.2(b). As these elements must also be orthogonal, they are set at right angles at some orientation to the diagonal. Without the presence of empirical data, a final choice of axes orientation is still possible. Obvious choices maximize the axis length and hence interval of representation for the related coordinate. Since the cross-section is convex, extrema from the figure's centroid lie at vertices. Other choices exhibiting high symmetry include the midpoint of any edge of the circuit which defines the hexagon. Note that quadrature of the basis element $\beta_2(\lambda)$ to $\beta_3(\lambda)$ causes both forms to appear. Methods of representation based upon geometric properties of zonotope reflectance solids (and not directly upon empirical data) are treated at greater length in Chapter 7 (Inflation).

Figure 5.3: (Un)modified Cube Basis Vectors

Algebraically, the Gram-Schmidt process may be applied to enforce the conditions of orthogonality. The QR method is its discrete analog in matrix form (Golub and Loan, 1989d). Here, a simplified reflectance space having three discrete samples per reflectance function ($n = 3$) forms a $3 \times 3$ matrix $\mathbf{B}$. A MATLAB session produces the orthogonal and symmetric *orthoprojector* matrix[7],

$$\mathbf{B} = \begin{bmatrix} .5774 & .5774 & .5774 \\ .5774 & -.7887 & .2113 \\ .5774 & .2113 & -.7887 \end{bmatrix}. \qquad [5.46]$$

The related colour cube is identical to the above except the coordinate system has undergone a rigid rotation. The related basis functions associated with the chosen axes are piecewise discrete analogs to the orthogonal set of continuous functions $(1, \sin, \cos)$ seen shortly. Such basis vectors are closely related to the Walsh functions (Hadamard basis). The set of cubic basis elements and their modified counterparts appear in Figure 5.3. The discrete nature of the reflectance functions ($n = 3$) has given rise to a circuit of six "chromatic" vertices which form a hexagon by action of the cutting plane. The edges form a *belt* traverse whose associated reflectance functions lie midway between the white and black apses of full and non-reflectance. The belt is not $C^1$ continuous yet is planar. (Belts are the subject of

[7] That is, $\mathbf{B} = \mathbf{B}^{\mathrm{T}}$ so $\mathbf{B}^{\mathrm{T}} = \mathbf{B}^{-1}$ (involution), hence $\mathbf{B}^2 = \mathbf{I}$.

Figure 5.4: Reflective and Additive Solids

§7.4). Figure 5.4 presents both (a) the halved reflectance solid and (b) plots of the (additive) chromaticity coordinates of the ten vertices of (a) under illumination by source E. A harmonization of both methods is possible because both methodologies employ linear methods. In particular, the chromaticity of hexagon's reflectances plots as a hexagon distorted by projection. More important, the vertices may be arranged as three pairs of two points, with each pair lying precisely along the edge of the triangular gamut whose three defining vertices are the projection of the RGB edges of the cube (not removed by the cross-sectioning). This condition can also be derived by geometric projection.

With larger $n$, application of the MSSM method creates discrete basis vectors $\mathbf{b}$ which better approximate continuous reflectance elements $\beta(\lambda)$. The continuity of such functions produces reflectance solids whose belt and general surface have a more sculpted shape. The $n$ dimensional model is thus encases a continuous surface, which may be created given continuous reflectance functions, as when analogs to the SVD operating on a space of continuous functions (Horn, 1950) is substituted. Such an example is presented below.

### 5.3.3 Orthogonal Sinusoids

At the other extreme, a continuous space spanned by three orthogonal functions may be used as a reflectance basis. Here $n$ may be taken to be arbitrarily large. The basis set

$$\begin{aligned} \bar{\beta}_1(\lambda) &= 1 \\ \tilde{\beta}_2(\lambda) &= \cos\theta \\ \tilde{\beta}_3(\lambda) &= \sin\theta \end{aligned} \qquad [5.47]$$

with $-\pi \leq \theta < \pi$ is orthogonal and provides representation of continuous reflectances with three degrees of freedom ($m = 3$). The reflectance solid belonging to this Fourier basis may be determined by solving for the conditions of extreme value:

$$\mathcal{R}^{n-cone} \equiv 0 \leq x_1(1) + x_2\cos\theta + x_3\sin\theta \leq 1. \qquad [5.48]$$

in which $\theta$ forms one period across the domain of wavelength.

Linear combination of the second and third elements will always produce a sinusoid of like frequency but having phase advance $\phi = tan^{-1}x_3/x_2$ and amplitude $A = \pm\sqrt{x_2^2 + x_3^2}$. This is the *phasor* method (Libbey, 1991) which regards the phase and amplitude as a polar representation of rectilinear components at quadrature.

Exploration of the chromaticity coordinates exhibited by reflectances having a unit period and arbitrary phase offset (the same basis but in polar form) were explored previously (D'Zmura and Lennie, 1986). That work extends a prior method (Buchsbaum and Gottshalk, 1983) which also used the Fourier basis $(1, \cos, \sin)$ but in the representation of spectral power distributions. Substitution into [5.48] gives

$$0 \leq x_1 \pm \sqrt{x_2^2 + x_3^2} \leq 1. \qquad [5.49]$$

For a given value $0 \leq x_1 \leq 1$ the second term contributes amount when $x_2 = \sqrt{1 - x_3^2}$. This may be parameterized instead as $x_2 = \cos(\phi)$, $x_3 = \sin(\phi)$ using $\phi$

as above. The locus of points $x_2$ and $x_3$ then describe a circular, coaxial disk. This disk is concentric and normal to the axis $\bar{\beta}_1(\lambda)$ because of orthogonality against its defining axes $\tilde{\beta}_2(\lambda)$ and $\tilde{\beta}_3(\lambda)$.

The radius of the disk is easily found as a function of $x_1$ by considering the simultaneous conditions $x_1 \pm A = 0$ and $x_1 \pm A = 1$ which are the limits of the inequality. This gives

$$A = \begin{cases} x_1 & \text{if } 0 \leq x_1 \leq \frac{1}{2} \\ 1 - x_1 & \text{if } \frac{1}{2} \leq x_1 \leq 1 \\ else & \text{undefined} \end{cases} , \qquad [5.50]$$

revealing a unit rhombus formed as the cross-section of any plane containing the axis $\bar{\beta}_1(\lambda)$. This radially-symmetric solid of rotation is the double-cone about the axis $x_1$, which introduces the symmetric white apex lacking in the solid previously appearing in Figure 4.6. The double-cone is presented in the full context of the interactive modeler, here in the main window (Figure 5.5). A rhombic cross-section appears in the subwindow titled SLICE. In the CIEXYZ subwindow the chromaticity plot of a select portion of the cone's belt vertices appears, reconstructing the plot appearing in a previous work (Buchsbaum and Gottshalk, 1984). The common figure is necessarily an ellipse, as it is the image of the circular disk derived above appearing under affine projection.

The apses of the cone determine points of black and white reflectance $\rho(\lambda) = 0$ and $\rho(\lambda) = 1$, respectively. As seen later, these are cusps because the limiting conditions of the inequality are met at more than $m = 3$ distinct wavelengths. The locus of points forming the belt correspond to reflectances having two extreme. The belt lies entirely within the plane $x_1 = \frac{1}{2}$. Substitution of this value reveals a set of sinusoidal reflectance functions identical save for a sliding peak wavelength $\phi$. Note that periodicity guarantees that when the maximum (minimum) wavelength occurs

Figure 5.5: Sinusoidal Bi-Cone

simultaneously at both extreme ends of the domain of defining wavelength. It is this continuity which removes any joints from the belt as seen with the cube. (It may be considered a dense set of joints in continuous distribution, as the direction of the belt traverse continually changes). These surface features are described further in Chapter 7.

Before leaving this example, it should be emphasized that the sinusoidal interrelation of coordinates $x_2$ and $x_3$ coincidentally matches the sinusoidal nature of their associated reflectance elements $\tilde{\beta}_2(\lambda)$ and $\tilde{\beta}_3(\lambda)$. This is a consequence of the identity $\sin^2\xi + \cos^2\xi = 1$ (valid for all $\xi$) and the stated limiting conditions. In general, traverses along the surface of the solid (defining an interrelation between the coordinates of the associated points) need not bear any resemblance to the basis functions which define their reflectance space.

### 5.3.4 Summary

The two cases above represent the limits of surface geometry for zonotopes. The double-cone exhibits a planar hue circuit having full $C^1$ continuity along its traverse; the cube lacks all these properties. Both share a central symmetry of inversion and an internal grey axis whose endpoints are cusps (degenerate vertices) defining the black and white ideal reflectances. Since the initial axis (basis element) encodes achromatic information on overall reflectance, the remaining axes (elements) encode hue information. In particular, the hexagonal facet of in Figure 5.4(a) is a surface of constant reflectance (constant luminance under the equipotential source). Note that the "hue" of the solid is defined without reference to any source of illumination.

### 5.3.5 Related Work

Reflectance spaces define surfaces possessing *colour constancy,* a valuable attribute and the subject of much on-going research within the related fields of computer vision (Maloney and Wandell, 1986), (Wandell, 1989), (D'Zmura, 1992) and in the *illumination invariant* biological models of trichromaticity (Yilmaz, 1962), (Brill, 1978), (Brill, 1979).

The *reflectance solid* unique to this thesis encodes colour information without regard to a source of illumination. It is an analog of the *illumination solid* described (and so named) in §3.2.1 which defines the set of all possible gamuts of a set of luminous sources. The reflectance solid generalizes the familiar *RGB colour cube* of computer graphics and related additive solids, described in §3.2.5. As with those solids, an achromatic axis may be established, the remaining axes then defining hue either along a continuous surface (CIEXYZ) or a hexagon (HSB, HLS), the latter a consequence of sectioning the cube along its body diagonal.

## 5.4 Conclusions

Modifications to the traditional method of orthogonal basis construction (SVD) form reflectance spaces whose associated reflectance solids exhibit a high degree of symmetry. The use of the $MSSM$ method is advocated based on metrical properties of its modified basis; comparison of (un)modified spaces by empirical study of gamuts is treated in the following chapter.

Symmetric reflectance space universally allows the exact representation of both the white reflectance $\rho(\lambda) = 1$ and the complementary reflectance functions $\rho'(\lambda) = 1 - \rho(\lambda)$. The associated reflectance solids possess central symmetry, relating them

to the zonotopes. The prototypical reflectance solids of symmetric reflectance space include the hypersphere, the hypercube and the bi-cone. The latter two are related to existing methods *a priori* orthogonal methods which employ either the diagonal or the Fourier bases.

# Chapter 6

# Deflation and Reflectance Matching

Deflation describes the mapping of an $n$-dimensional linear reflectance space onto an $m$-dimensional subspace, where $n > m$. Because any point in reflectance space defines a unique reflectance function, deflation maps one or more reflectance functions onto a single reflectance function on the subspace, thereby creating an approximation. Deflation finds use in representing and manipulating descriptions of object surfaces at high precision by using reflectance coordinates having a reduced number of components $m$. This reduction is not merely desirable: it is often a constraint of the problem specifics, as here. Deflationary methods are well-suited to solving various problems of digital colour, including the specification of illumination-independent colour and the device-independent control of a digital printer.

Throughout this chapter, methods of deflation are examined and refined in relation to a new, central problem termed *reflectance matching:*

> *Given a coloured surface original with known reflectance function, produce a matching facsimile whose reflectance function lies in (arbitrarily) close agreement.*

A match produced in this fashion does not exhibit metamerism under a change of source of illumination[1]. This is the fundamental attribute of reflectance matching which makes a viable solution desirable.

The reflectance match is typically a colour hardcopy *facsimile* produced under digital control. In contrast, the original need not be a physical artifact but could be a surface synthetically defined using CAD/CAM or graphic rendering software. At the other extreme, when the original is itself a printed image, reflectance matching offers a new means of solution to the long-standing, difficult problem of reprographic colour correction (Stone, Cowan and Beatty, 1988). Solution based on spectral reflectance matching represents a sharp departure from traditional methods of colour correction (Brown, 1955), which consider the matching of (additive) chromaticity coordinates of media characterized under a well-defined source of illumination, hence most often by a linear $(3 \times 3)$ change of basis with possible second-order correction.

The chapter is in five sections. The first is an overview of analytic deflation which establishes the practical techniques used. The second is an overview of traditional methods of colour correction. Taken together, these define the experimental

[1]Metamerism of reflectances in other contexts has been explored elsewhere (Takahama and Nayatani, 1972), (Burns, Cohen and Kuznetsov, 1989), as has geometrical metamerism (Cohen and Kappauf, 1985).

methods of deflation which are explored in the remaining three sections of the chapter. These are presented in order of successive refinement. The third section applies the "first principles" methods of deflation using the SVD. The liabilities listed in §4.2.10 are revealed in the course of study. In section four, the modified methods of §5.1.5 are retrofitted and comparative results provided. (The exposition of both sections are in close parallel). Successful reflectance matches are made to coloured paper stock included in the binding of this document, allowing the thesis manuscript *per se* to serve as an original, with facsimile presented along side for direct visual comparison. The third section employs multiple reflectance spaces to gain greater accuracy in solving a more traditional reprographics problem, which is subject to the constraint $m = 3$. This is also an instance of problem solving using multiple dissimilar spaces, an area of on-going study. The chapter closes with a brief summary.

## 6.1 Digital Printing: Overviews

### 6.1.1 Deflation Models of Printer Reflectances

Deflation of an $n$ onto an $m$-dimensional linear reflectance space, or $\mathbf{B}^{(n)} \overset{D}{\rightarrow} \mathbf{B}^{(m)}$, may be considered a rigid rotation of the frame of reference of the $n$-space followed by an orthogonal projection onto $m$-space, removing $n - m$ of basis elements. The projection of any reflectance coordinate $x^{(n)}$ onto its image $x'^{(m)}$ establishes the point of closest approach in the subspace (Apostol, 1969e). Formally, the mapping of a reflectance function $\mathbf{r}$ in $n$-space (here $\mathbf{r}$ is isometric with its reflectance coordinates by [4.6]) onto the deflated reflectance coordinates $\mathbf{x}'$ in $m$-space may be

easily expressed using the matrix notation

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_m \end{bmatrix}^{\mathrm{T}} = diag([1,\ldots,1,0,\ldots,0]) \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \\ \vdots & \vdots & & \vdots \\ & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad [6.1]$$

indicating (from the right) that $\mathbf{r}$ is transformed by an orthogonal change of basis $\mathbf{T}$ followed by a identity $\mathbf{D}$ having $n-m$ trailing ones along the diagonal. The action of this matrix provides conformality while eliding the elements not present in the subspace $\mathbf{B}^{(\mathrm{m})}$ and which define the null space of the deflationary transformation. When the change of basis is created by regression methods such as SVD, these trailing diagonal elements decrease as $\mathbf{D}$ is (nearly) singular; the loss of representation is then small. Reassociating $\mathbf{D}$ with the transformation matrix forms the familiar rectangular change of basis matrix $\mathbf{B}^{(\mathrm{m})}$ used (together with its transpose) to map between reflectance coordinates [4.3] and reflectance functions [4.5].

### Null Space and Black Reflectance

The columns lost in forming $\mathbf{B}$ define the null space of the deflation. These elements span the space of *reflectance metamers* of $\mathbf{B}^{(\mathrm{m})}$. That is, each $j$th elided basis vector $\mathbf{b}_{j>m}$ defines a model reflectance function which may be added in linear combination to any reflectance vector, thereby perturbing the underlying reflectance function without changing its deflated reflectance coordinates, since

$$\tilde{\mathbf{x}}^{(\mathrm{r}+\mathrm{b_j})} = \mathbf{B}(\mathbf{r}+\mathbf{b}_j) = \mathbf{B}\mathbf{r} + \mathbf{B}\mathbf{b}_j = \mathbf{B}\mathbf{r} = \tilde{\mathbf{x}}^{(\mathrm{r})} \qquad [6.2]$$

whenever $\mathbf{B}\mathbf{b}_j = 0$, which holds by definition for $m < j \leq n$. Geometrically, these columns span the space of points having a common point of projection within

the deflated space. These points define reflectance metamers by analogy to the metamerism of light possible in the perceptual space of additive colour.

Metamerism of light in relation to surfaces may be modeled by introducing a source of illumination (and by considering perceptual response). Specifically, a three-dimensional reflectance space can be created in which the elided null space defines all reflectance metamers. Pioneering work in the field (Wyszecki, 1958) considers such linearly independent reflectance functions *metameric black* reflectances. Each necessarily has tristimulus values $X = Y = Z = 0$ under the chosen source of illumination. A set of twenty-seven model reflectances which are black under CIE source C are tabulated by (Wyszecki and Styles, 1982d). The profile of each is everywhere zero save for three narrow (1nm) "tines" of fixed wavelength and tabulated height; the wavelength of the fourth tine of unit height varies with curve number. Wyszecki's early and insightful method of approach provides significant step toward the formalization of linear reflectance space[2]. First, physical reflectances may be decomposed into sums of reflectance components (Eq. 3(3.8.2)). Second, the latter are abstractions which need not satisfy $0 \leq \rho(\lambda) \leq 1$ (Eq. 4(3.8.2)). Finally,

> A mathematically more elegant basic set of spectral reflectance functions of metameric blacks may be derived by orthogonalizing the functions (*ibid,* p. 188).

The retabulated functions, when plotted, reveal a continuous nature. The formalization of these principles in the previous chapter help define Euclidean reflectance spaces. The concepts of metameric black were extended during the course of this research by colleagues (Brainard, Wandell and Cowan, 1989).

[2]Wyzecki dedicates thirty-seven pages to "Metameric Color Stimuli" in §3.8 of his *magnus opus.*

Deflation techniques advances Wyszecki's method of approach, for no source of illumination is presupposed: operations are strictly upon reflectance functions. Here, the basis elements lost through deflation must approximate metameric black reflectances for an unspecified set of illuminating sources. Thus, a representative set of basis vectors must be created merely through analytical insight or empirical observation with reference to the broad class of reflectances whose approximation is anticipated.

### Deflation Requirements

Deflation is fully defined by the $m \times n$ matrix which defines the orthogonal basis of the new reflectance space. Successful deflation locates a suitable axis of projection (following rigid rotation) whose removal minimizes loss of representation accuracy. Three properties help determine the nature of the basis elements in a "best" deflation algorithm. Each consider deflation onto a $m$-space which in the limit approaches $n$, that is, deflations which do increasingly less work. Useful properties and terminology are set forth below.

The property of *exact convergence* guarantees that as the dimension $m$ is increased to the limit $n$, the deflated spaces of increasing size converge to the original. This is guaranteed provided that the parent $n \times n$ matrix is independent (as with the tine basis), thereby having full rank. Notice that two $m$-dimensional deflations $\mathbf{B}^{(n)} \xrightarrow{D} \mathbf{B}^{(m)}$ and $\mathbf{B}^{(n)} \xrightarrow{D} \mathbf{B}^{(m')}$ need *not* span the same subspace. This is self-evident when the bases are produced by different orthogonalization methods, subtle when a single technique is applied to a distinct sets of reflectance gamuts. Regardless, the inflation of any $m$-space ultimately converges at a common $n$-space, whereby $\mathbf{r} = \mathbf{I}\ \mathbf{x}$.

Second, the property of *monotonic convergence* assures that deflation onto spaces having progressively larger dimension $m$ must must increase the representational accuracy of smaller spaces. (This condition is sufficient but not necessary for the previous property to hold). A stronger form, here termed *unique convergence* requires a full ordering of the basis elements be defined. This assures that two distinct deflationary sequences

$$\begin{aligned} \mathbf{B}^{(\mathrm{n})} &\overset{\mathrm{D}}{\rightarrow} \mathbf{B}^{(\mathrm{i})} \overset{\mathrm{D}}{\rightarrow} \mathbf{B}^{(\mathrm{m})} \qquad n < i < m \\ \mathbf{B}^{(\mathrm{n})} &\overset{\mathrm{D}}{\rightarrow} \mathbf{B}^{(\mathrm{j})} \overset{\mathrm{D}}{\rightarrow} \mathbf{B}^{(\mathrm{m})} \qquad n < j < m \end{aligned} \tag{6.3}$$

passing through distinct ($i \neq j$) intermediate bases nonetheless arrive at the common basis $\mathbf{B}^{(\mathrm{m})}$, thereby defining a canonical path. The SVD and MSSM orthogonalization techniques explored thus far provide an ordering of basis elements and thus satisfy this property when the first $m$ elements of any orthogonal basis are retained.

### 6.1.2 Traditional Printing (Colour Correction)

Traditionally, the process of matching a print to a film original is called *colour correction.* This step is required because system non-linearities make direct control of the reflectance (transmission) functions impossible. Photographic effects (Mitchell, 1984) include the reduction of contrast through lens flare or spatial variance in the modulation transfer function (image response to line pairs at various spacings). Ink is less well-behaved. Underlying non-linear processes include the non-uniform transfer of small ink particles to the stock, the scattering effects between ink layers and (perhaps worst of all) the discontinuous nature of the interface between adjacent ink layers (Yule, 1967b).

### Printer Behaviour and Correction

The mapping from input prescription to output colour defines the *transfer function* of the printer. An accurate model of this forward function allows the behaviour of the printer to be predicted under any input conditions. Inverting this mapping provides the means of correction: for any desired output the required inputs are calculated. The latter are the *printer coordinates* of the device. Correction therefore concerns the mapping which interrelates two vector spaces. Were the mapping linear and the device viewed under a single source of illumination, a $3 \times 3$ linear transformation matrix would suffice, as with the CRT model. This approach is often taken on naïve systems which assume global linearity of the print function.

Numerous methods exist to model the printer function. Traditional models such as the Neugebauer equations (Neugebauer, 1937) make simple physical assumptions. These allow solution by back substitution of the partially coupled non-linear terms. Similar simplifications allow for *mask making* (Pollak, 1955), in which photographic processes allow for massively parallel optical (analog) computation.

Yule suggested in almost visionary fashion (Yule, 1967a) that techniques of linear interpolation would prove superior to this task and become practical with the advent of high-speed digital computation. The sampling approach he advocated represents an emerging method of printer characterization. Related methods were first explored (Clapper, 1961) using a twenty-seven parameter model[3] as an alternative to the Neugebauer equations. Yule's prediction is based on the observation that the printer's transfer function is continuous and thus linear in the neighborhood of the sampled points, as shown by Taylor expansion. This approach considers the accuracy of a print model better gauged by its quantitative information (the

[3]That is, a $3 \times 3 \times 3$ gamut is sampled.

number of underlying parameters) and not its qualitative attributes (the model's ability to characterize the printer's non-linear physical behaviour). Interpolative methods of colour correction of CMYK and/or RGB coordinates in the additive domain are of emerging importance in both the industrial products (MacDonald-Dettwiler and Associates, 1988) and the research domain (Stone and Wallace, 1991) and are reviewed elsewhere (Paeth, 1989a), (Paeth, 1989b).

### Tabular Correction Methods

The automated matching of a digital image requires a model of the behaviour of the printer. This may be established by studying the printer's gamut. The digital printer produces a finite number of surfaces under control of a finite number of printer channels having discrete levels. (The Kodak SVC-6500 has three channels of sixty-four levels each, yielding a print gamut of $262,144$ colours). Being too large to characterize in practice, the gamut may be sampled instead at regular intervals. For instance, set of print values $\{0,\ 21,\ 42,\ 63\}$ taken as a Cartesian product on all three channels provides a $4 \times 4 \times 4$ set of printer values which represents the full range of the printer's gamut, albeit at a coarse sampling. Geometrically, the sampling coordinates may be represented in a space whose axes define the values presented to each input channel on the printer, seen in Figure 6.1.

The printer generates (in the sense indicated by the upward pointing arrows appearing in Figure 2.1) a set of reflective surfaces given this set of input coordinates. Production is economized in printing reflective surfaces in parallel: a *mosaic* of colour *tiles* is created on one or more output sheets. These form a discrete sampling of the printer's gamut. The mosaic produced by the $4 \times 4 \times 4$ gamut appears in Figure 6.2 as a physical artifact. (Note: the thermal sublimation print is not present in most bound copies of this thesis. Instead, the underlying encapsulated

Figure 6.1: Printer RGB Input Lattice

Postscript document reproduces the gamut using the "colorimage" extension to encode the four RGB print densities by using byte values {0, 85, 170, 255}, allowing automatic substitution of colour folios when reproduced upon a digital colour printer.)

Traditional methods estimate the behaviour of the printer by recording colorimetric values (such as tristimulus values or CMY print densities) for a small number of samples under a standard source of illumination, traditionally D5000. The values of the *test strip* serve as parameters to a simplified, partially-coupled model which may be easily inverted, thereby predicting the printer's inverse *transfer function* in discrete form. This functions computes the printer coordinates required which produce colour output (print coordinates) having a desired response, such as XYZ values or reflectance coordinates (see the following subsection).

The CIEXYZ chromaticity coordinates $(x', y')$ of all sixty-four samples are plotted in Figure 6.3. These may be considered an unusually large test strip of the printer data. The three digit plotting symbol indicates the printer coordinates

Figure 6.2: Kodak 8x8 Mosaic (4x4x4 Gamut)

used to create the sample. The numerals $n = 0 \ldots 3$ appear in RGB order with $85n$ giving the byte value of that channel (000-black, 300-full red, etc.), with the gamut's corners printed in two letter abbreviations, such as "Ye" for yellow. The resulting space figure is the *illumination gamut* of the printer within tristimulus space under source E.

In three dimensions, XYZ values of the SVC-6500 gamut may be plotted in XYZ tristimulus space, as in Figure 6.4. This reemphasizes the concept of a print gamut as a geometric figure within a colour space. Rendered in perspective, the $XYZ$ colorimetric values were printed as points along the corresponding $x$, $y$ and $z$ coordinate axes. This wire-frame figure may be regarded as the non-linear transfer of the uniform cubic lattice presented in Figure 6.1. This suggests interconnecting points having neighboring printer coordinates along the output domain, better

Figure 6.3: Printer Chromaticity Gamut

Figure 6.4: Kodak Tristimulus Gamut (3D perspective)

revealing the forward transfer function.

Consistent with the current research (Stone, Beatty and Cowan, 1986) is the pronounced non-linear "toe" near the black point. Because black is formed by superimposing all three colorants, approaching black means a greater build-up of layers. Surface gloss and scattering effects then work against the creation of a darker colour. When surfaces are glossy and not matte the apex may even "double-back": full ink coverage may not provide the print having greatest print density (least reflectance). The geometric gamut is no longer convex. Moreover, the mapping is no longer a one-to-one function. It is the presence of such conditions which complicate the task of locating a suitable table entry when computing the inverse map.

The 3-D perspective rendering is superior in revealing both the coupling *and*

Figure 6.5: Kodak Tristimulus Gamut (2D projections)

the non-linearity near the black point. By way of comparison, the "front", "side" and "top" orthographic view of the solid suggest merely the former. Projections onto the $z$ and $x$ planes appear in Figure 6.5.

The obvious next step is the substitution of reflectance coordinates of each tile for the tristimulus values while reusing methods of trilinear interpolation. This is described in the following sections.

### Toward Deflation-based Correction

Deflationary print matching records instead the reflectance coordinates associated with each each tile. Generally, the transfer function is linear for small perturbations from a known point of characterization. By recording both the discrete input and output coordinates as a tuple, an accurate model of the *reflectance transfer function* is created. That is, given any RGB input (as for the Kodak printer), a set of table

entries may be interpolated which describe the reflectance of a print anticipated under these conditions.

When the dimension of the deflationary space matches the input space ($m = 3$), the inverse print function may be computed merely by transposing domain and range components of each tuple and reapplying identical methods. That is, given a set of reflectance coordinates lying in the space of characterization, a set of print values can be estimated. Reflectance coordinates are easily created given a reflectance vector using [4.5]. This creates a means of mapping the appearance of an object's surface (whose reflectance may be characterized using a spectroradiometer) onto a facsimile. A unique aspect of this method is that the match takes place without reference to any source of illumination.

A series of experiments in reflectance matching using a three-dimensional reflectance space and methods of tri-linear interpolation are described in the following three sections.

## 6.2 Print Matching – Procedural Overview

Reflectance matching of physical samples employs three steps: characterization, representation and reproduction. The first step quantifies objects' surfaces in terms of colorimetric information. The second step maps this information into a set of *print coordinates.* The final step produces the digital facsimile whose surface properties are estimated to approximate the original. (When desirable, this artifact may be resampled to provide a *closed-loop* check upon the methods). The three steps are described below.

### 6.2.1 Materials Characterization

Experimentation throughout this work is underpinned by data produced on a custom spectroradiometer. It provides a full disclosure of any surface's reflectance function, in contrast contrast to the *colorimeters* and *colour densitometers* of traditional reprographics. However, custom software which maps reflectances onto chromaticity and related trilinear coordinate systems (Appendix E) provided "virtual" instruments of this type, as used in the graphing and annotation of Figure 3.3 and elsewhere. A computer controlled motion stage which carries reflective or transmissive samples allows the unattended characterization of large mosaics.

Initial work was performed on a calibrated Perkin-Elmer unit at the Canadian National Research Council in Ottawa. Subsequent work at the CNRC was conducted on a custom prototype which recorded a $16 \times 16 \times 16$ gamut recorded as a $64 \times 64$ mosaic reproduced on 8.5" $\times$ 11" Ektachrome 5017 sheet film[4]. Multiple spectra were averaged to improve amplifier signal-to-noise ratio. Each mosaic tile was sampled at three spatial locations and the sample having median total output retained as representative. This step removed the influence of any surface irregularity or lens flare. Treatment of calibration, data compression and exchange on sixty-four 5.25" PC floppy disks is treated in the second appendix of (Paeth, 1989a). The data is presented graphically in §8.1.

A uniform file format (in ASCII) was created which presented data as column vectors. This choice allows the representation of discrete reflectance vectors (or reflectance basis matrices) as files having one (or a few) columns and an arbitrary number of rows, thus avoiding software limitations in file line length.

[4]The author wishes to thank Sean Adkins and John Keightley of Pthalo Systems, Langley B.C. for assistance and materials.

A custom spectroradiometery system was eventually placed in-house. It borrowed in large part on the experience and materials of its predecessor. The software migrated from a PC into Macintosh-based control. This provided network access, greatly increasing turn-around on gamut jobs. Output detail is stored in a one-file-per-tile basis with file name derived from sheet, row and column number. This allows the data files to exist amicably within and across most commonly encountered file systems.

The reflectance spectra of the "red" tile of the SVC6500 appears in Figure 6.6. In was created given an $RGB$ input triple $(255, 0, 0)$. The output has been normalized by dividing each sample wavelength against the sample of corresponding wavelength from a white reference tile (tool WHITENORM). This provides relative reflectance independent of illumination source and reflectance function of the paper stock as by [2.2].

The companion graph accompanies the excerpted reflectance table. It was also created under automatic control using the LEMMING interactive drafting system (Paeth, 1988) which produced the line art appearing throughout this document.

### 6.2.2 Coordinate Mapping

The second step is provided by methods of deflation described throughout this chapter, augmented by recent techniques in *gamut mapping.* Here, the behaviour of the printer was probed by producing and characterizing a *reflectance gamut* of samples, whose reflectances were mapped onto unique reflectance coordinates. These are associated with the three control inputs of the printer under study (RGB values on the domain range $[0 \ldots 255]$ for the Kodak SVC-6500). This one-to-one mapping provides an approximating reflectance function for given RGB values

Figure 6.6: Transmission Spectrum (Ektachrome 5017)

at regularly spaced points along this Cartesian lattice. Tri-linear interpolation software provides a means of mapping those printer coordinates lying at fractional distance within the interval defined by the eight vertices of the unit bracketing cube.

The software providing the non-linear mapping between (three) reflectance coordinates and (three) print values closely resembles conventional colour correction software. Traditional systems map between additive colour spaces (RGB, XYZ or LUV) whose components are non-negative. Reflectance coordinates do not have this restriction. Removing any range or domain restrictions thus generalize the conventional methods. This proper treatment was set forth as a design requirement by the author for general gamut mapping software within CGL in anticipation of this reuse with reflectance-based data. The evolution of specific requirements together

with the high turnover of student programmers drove the creation of a custom trilinear interpolation routine INVERTRGB. The tool provides a simple means of inverse mapping: a program switch reverse the input and output coordinates of the table presented it. For more details see Appendix C.

A drawback with table interpolation methods is that they are potentially non-invertible. The table search must account for unsolvable conditions. This arise when solution of the given reflectance lies outside the printer gamut (as when matching highly saturated colours) or when multiple solutions are possible, which may occur with very dense printing combinations. These difficulties can plague the (additive) print matching of reflective media to CRT-based luminous colours. Fortunately, the gamut of the reflective printer is larger than the gamut of reflective colours commonly encountered.

### 6.2.3 Printing

Finally, the printer values drive the digitally-controlled printer[5] yielding a final artifact. For reasons of simplicity and experimental consistency, the Kodak SVD-6500 was installed (Appendix D) set up "out of the box" as no adjustments to the hardware look-up table or contrast controls were made. However, the first model was returned to Kodak for a new print head and general adjustment given minor non-uniformity.

This printer's capabilities are summarized briefly. The Kodak SVC-6500[6] stores a $512 \times 512$ pixel array of $R$, $G$ and $B$ byte planes loaded by RGB video or composite NTSC video signal or optional Centronics parallel interface. Imaging is by

[5]Nearly all commercial printmaking is now under digital control.

[6]The first shipped in Canada, it has since been superseded by the model SVC6510.

wax transfer (thermal sublimation) onto 3.5" × 4.5" specially treated stock using a thermal process. The imaging head provides sixty-four discrete heat levels which sublimates cyan, magenta or yellow fields appearing in cyclic order on a non-reusable film ribbon. The print gamut thus contains $(2^6)^3$ or 262,144 colours. A custom RS-232 to Centronics interface (Appendix E) and related software allowed computer control of the device to 19,200 baud, which is 90% of its maximum transfer rate.

The gamuts used to characterize the SVD-6500's overall (forward) print function were produced using the a custom tool IMMOSAIC (Paeth, 1986), (Paeth, 1987a). The tool created multiple sheets of output containing tile mosaics arranged in a square pattern. Tile positions were permuted using a deterministic pseudorandom number generator. This step further reduces errors introduced by colour spatial dependency.

Companion output presents each the sheet, row and column position of each tile by line, with annotation included for tiles lying at the corner of the colour cube. The detail is in Unix command syntax, allowing the description file to act as a command script for actions taken against each tile comprising the mosaic. The file is also useful in recording the permutations of the mosaic's tiles and in identifying the corners of the gamut's colour cube, as these entries are flagged with text annotations. Eleven lines of detail are excerpted below. These correspond to the $4 \times 4 \times 4$ gamut (artifact) encoded as an $8 \times 8$ mosaic which appeared previously in Figure 6.2.

```
#!/bin/csh -f
# 1 sheets of 8x8
# with 64 samples per sheet
```

```
# and 64 samples in total
#    ShRwCl RGB (label)
#
./op 010101 131
./op 010102 130
./op 010103 313
./op 010104 101
./op 010105 301
./op 010106 012
./op 010107 103
./op 010108 003 blue
./op 010201 300 red
   .
   .
   .
./op 010807 030 green
./op 010808 211
```

The culled file is helpful in locating specific tiles such as the reference white used to normalize the reflectance data. Note that the table is unnecessary for methods of global analysis such as SVD, whose characteristic output vectors are invariant with respect to permutation of the columns (reflectance functions).

## 6.3 Print Matching – Global SVD Method

The three-fold goal of this section is to demonstrate the overall procedure of reflectance mapping, to estimate the goodness of fit of the method and to provide artifacts for visual and analytic comparison. First SVD is applied to a set of mosaic tiles (a $4 \times 4 \times 4$ gamut sampling) to provide a ranked set of orthogonal basis (reflectance) functions. These define a nested sequence of deflationary spaces. Next, the tile reflectances are reexpressed under the new system and their goodness of fit computed. Finally, reflectances of coloured stock are also measured and reexpressed

in three-dimensional reflectance space and ultimately mapped onto printer values (and facsimiles struck, (Paeth, 1989a)). The conclusions serve as impetus to the sections which follow.

### 6.3.1 Gamut Production

The $4 \times 4 \times 4$ input gamut appearing in Figure 6.1 was produced and imaged on the SVC6500 printer. The spectra for the eight corners of the cube are presented in Figure 6.7. Chromaticities under CIE source E (the equipotential source) of the vertices annotated each chart (these were plotted in Figure 6.2).

Note that the white sample is the ideal reflectance $\rho(\lambda) = 1$ because of normalization against the source of illumination. As a consequence, its chromaticity lies at $x' = x' = \frac{1}{3}$ when illuminated under source E. The $(x', y')$ chromaticities forming a conventional (additive) printer gamut were plotted in Figure 6.3. Note that the eight corners of the printer gamut (themselves a convex hull) do *not* form a convex set when projected using conventional models (Figures 3.3 and 6.4).

### 6.3.2 SVD Basis Determination

A set of basis coordinates was derived from the tile reflectances by the singular value decomposition. These are the eigenvalues of the covariance matrix $\mathbf{S} = \mathbf{D}\mathbf{D}^{\mathrm{T}}$ where $\mathbf{D}$ is the $512 \times 64$ data matrix of columnar samples. Note that $\mathbf{S}$ here is larger than $\mathbf{D}$. (The opposite holds when analysing data sets of more than 512 samples). The custom tool ORTBASIS computed the element $(i, j)$ of the covariance matrix by dotting the $i$th and $j$th columns of $\mathbf{D}$. This is performed dynamically, reducing storage requirements at the expense of computation. Because $\mathbf{S}$ is symmetric only

Figure 6.7: Reflectance Spectrum (SVC6500 Printer)

half its elements are computed; the underlying custom matrix software makes this invisible by providing a uniform access mechanism. Finally, the eigenvectors having largest matching eigenvalue were removed iteratively from $\mathbf{S}$ using the *Power* (Hotelling) method. Since $\mathbf{S}$ is computed dynamically the residuals must also be removed on a per-access basis. This caused the computation cost to grow quadratically with each eigenvector. The rate of increase in cost was more than offset by the requirement for a very small percentage of vectors (typically eight) in a partial decomposition. The results were found to agree with full decompositions provided by MATLAB to within a change in sign. However, limitations[7] since removed in MATLAB matrix size required that its input data be culled to $170 \times 64$. This was done by retaining data at every third wavelength. (The size limitations in MATLAB motivated the creation of the custom matrix software). Computational methods and tool specifics are treated at greater length in Appendices A and E, respectively.

Plots of the first six reflectance vectors appear in Figures 6.8(a) and (b). Each vector is is annotated with its matching eigenvalue. These indicate the basis element's relative contribution in providing a global fit of the data set.

The weights associated with the first three eigenvectors are large. This is not surprising, as the printer has three independent print colours. The large fourth eigenvector indicates that much of the printer non-linearity is captured by the addition of a fourth eigenvector. Because SVD is a global method it is not intuitively clear what effect of increasing dimension has on the accuracy of any specific reflectance. Plots of chromaticity migration as a function of dimensional change (appearing in §6.4.7) shed further light.

[7] The latest version supports matrices of arbitrary size, but is a proprietary product. See also Chapter 8.

Figure 6.8: SVD Reflectance Vectors

**Dimensional Goodness of Fit**

The SVD ranks the basis eigenvectors in order of corresponding eigenvalue. Only an initial subset of basis elements need be retained, indicated by the eigenvalues and a suitable criterion. Statistical use often substitutes the square-root of each eigenvalue. This is justified because the SVD of a diagonal matrix $\mathbf{D} = (\mathbf{d}_1 \ldots \mathbf{d}_n)$ (a fully decoupled system) will reconstruct the values of $\mathbf{D}$ (in sparse form) after the square-root is performed. This follows because symmetric matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ becomes the diagonal matrix $\Lambda = (\mathbf{d}_1^2 \ldots \mathbf{d}_N^2)$ and the eigenvectors of the diagonal matrix is the diagonal. (Preference is divided among statisticians for the same reasons that $\sigma^2$ squared variance is often employed instead of standard deviation $\sigma$). For any real matrix $\mathbf{D}$ the covariance matrix $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ is semi-definite, so the square-root is never imaginary. This follows because a symmetric matrix is semi-definite *iff* it can be decomposed into the quadratic form $\mathbf{D}\mathbf{D}^{\mathrm{T}}$ as by the Cholesky decomposition (Golub and Loan, 1989b).

Traditional interpretations of SVD-based analysis offer two stopping conditions. Either the ratio of consecutive confidence values (the square-root of the eigenvalue) falls below a predetermined threshold (Tatsuoka, 1988), or the confidence value falls below unity (Hotelling, 1933). The second method is employed here, making the previous considerations moot as the conditions $\sqrt{\mu} \leq 1$ and $\mu \leq 1$ are identical as $\mu \geq 0$.

### 6.3.3 RGB to Reflectance Function

By either criterion, the reflectance gamut may be well fit by retaining the five principal components. The mapping is recorded as a sixty-four 8-tuples which associate the $(R, G, B)$ printer coordinates with reflectance coordinates $\mathbf{x} = (x_1, \ldots, x_8)$. Tri-

linear interpolation of the discrete map creates a continuous map. Thus, a compact and accurate prediction of the reflectances generated by the printer for any input is created. That is, the mapping

$$\mathbf{p}_{rgb} \rightarrow \mathbf{x}^{(\mathrm{svd})} \qquad [6.4]$$

is created, a useful goal. However, the form is not easily inverted because of dimensional mismatch between the input and output coordinates which form the 8-tuple. This is a liability: reflectance matching seeks printer coordinates given reflectance coordinates.

The first-order solution is to limit the number of reflectance coordinates to three, thereby providing a dimensional match. Methods of both forward and inverse mapping are then easily described and may be used interchangeably: given an algorithm which defines a continuous complete mapping given a set of discrete 3-tuple pairs will computed the inverse map if the defining tuples are swapped.

### Rendering the SVD Solid

A hypothetical surface whose reflectance is defined by three reflectance coordinates may be viewed on a CRT under some source of illumination $\Phi\lambda$. Here the scaled contribution of each reflectance element may be reassociated as

$$[\mathbf{x}_i\ \beta_i(\lambda)]\ \Phi_\lambda = \mathbf{x}_i\ [\beta_i(\lambda)\ \Phi_\lambda], \qquad [6.5]$$

indicating that each component of the illuminated surface (left-hand side) may be considered a set of radiant sources (right-hand side). These are under linear control as by the equations of generation [2.9] and superposition [2.19]. (This imparts some "physical" meaning to the representation of reflectances, not as products but as linear sums).

Mapping of illumination coordinates onto tristimulus values or RGB pixel coordinates is straightforward exercise and was conducted in §3.1.2 (see also [6.9], [6.10] and Appendix F). As before, linearity allows the association of all intermediate products into a single $3 \times 3$ transformation matrix. This is presented below

$$\begin{bmatrix} p_R \\ p_G \\ p_B \end{bmatrix} = \begin{bmatrix} .0572 & .0166 & .0583 \\ .0473 & -.0332 & .0566 \\ .0687 & -.0543 & -.0909 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad [6.6]$$

together with its inverse

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 13.04 & -3.549 & 6.151 \\ 17.54 & -19.69 & -1.017 \\ -.6260 & 9.092 & -5.745 \end{bmatrix} \begin{bmatrix} p_R \\ p_G \\ p_B \end{bmatrix} \qquad [6.7]$$

in which the phosphor chromaticities for the Silicon Graphics/Iris monitor are used. Note that the mapping is by coordinates: expressing the mapping as a change of reference frame between the illumination spaces having bases $\beta$ and $\Phi_\lambda^{(\text{phosphor})}$ are not possible because these continuous functions cannot be interrelated, though there metameric appearance can be. Note in particular that the point $\mathbf{x} = (1, 0, 0)$ along the $\rho_1$ axis produces nearly $RGB$ equal drive values as the first column of the first matrix is then selected. That is, under source E (a reasonable approximation for daylight), the unit grey surface calls for a pixel having near equal drive, as would be expected on a device having proper grey balance.

### Hardware Rendering and Linearity

The rendering task is simplified by the Iris hardware which provides linear interpolation of polygon colour based on vertex colour. Since the reflectance functions show linear change with respect to their world coordinates, the pixel colours of the

illuminated surface likewise change in linear fashion. Put another way, if the reflectances $\rho_\lambda^{(a)}$ and $\rho_\lambda^{(b)}$ associated with two points in space have pixel (illumination) coordinates $\mathbf{p}^{(a)}$ and $\mathbf{p}^{(b)}$ by premultiplication by transformation matrix $\mathbf{T}$, then the colour of a point at fractional distance $0 \leq \alpha \leq 1$ on the segment $\overline{AB}$ is

$$\alpha\mathbf{p}^{(b)} + (1-\alpha)\mathbf{p}^{(a)} = \mathbf{T}[\alpha\rho^{(b)} + (1-\alpha)\rho^{(a)}] \qquad [6.8]$$

for any intermediate reflectance, as is clear by linearity of forms throughout. (See also Appendix H.)

Thus, hardware pixel fill of scan lines or polygons using linear interpolation creates intermediate pixels having the proper value. There remains one fine point. The left-hand side of [6.8] represents a point interpolated in world-coordinates; the right-hand side in pixel (screen) coordinates. Accordingly, the spatial correspondence is not exact for polygons showing strongly "forced" perspective. That is, the midpoint of a linear segment (in world coordinates) does not in general correspond to the midpoint of its projected length (in screen coordinates). A parametric form used to interrelate points along a segment in screen coordinates is well-known (Newman and Sproull, 1971); the related bi-linear forms are common to methods of texture mapping (Kreiger, 1984). The specific problem of perspective in relation to the linear interpolation of pixel values appears elsewhere (Lansdale, 1991).

### The SVD Reflectance Solid

The reflectance solid created by retaining the three initial reflectance vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ may be rendered in Euclidean 3-space. The solid is the locus of reflectance coordinates having physically valid reflectances under the SVD basis. Ideally, the reflectance mapping of arbitrary samples should project onto metameric points in this space which are contained by the solid.

Figure 6.9: SVD Reflectance Solid

The custom modeling software tool META4D displays the solid in perspective on the SGI/Iris workstation, illuminated by (hypothetical) CIE source E. As before, faces are rendered under manufacturer supplied CIE $(x', y')$ chromaticity coordinates for the CRT. The hardcopy appears in Figure 6.9 and is a snapshot of the screen output (identical to Figure 4.3) but here reproduced[8] on SVC6500 media.

The screen image is therefore self-referential, as the abstract surface of the reflectance solid is rendered in its true colour, establishing a correspondence between its spatial location, reflectance coordinates and CRT phosphor drive levels under the chosen source of illumination.

As discussed in §4.2.7, the SVD-derived solid is "snubbed" at the white end as the ideal reflectance lies out of gamut. This suggests that the SVD basis is unsuitable for approximations of near-white surface. The approximation of the ideal white (that is, its the point of projection) lies beyond the face; its reflectance function appears in Figure 6.13(h) in §6.3.4. An in-gamut white approximation would lie on the reflectance solid's surface near the point of its penetration by the $\tilde{\beta}_1(\lambda)$ axis.

[8]The same caveats concerning Figure 6.2 apply here.

Figure 6.10: SVD Reflectance Gamut (3-D perspective)

**SVD Gamut and Reflectance Coordinates**

The ranges of the reflectance coordinates may be explored directly by plotting the reflectance coordinates of each tile sample in 3-space (Figure 6.10). This 3-space lattice is the projection of the reflectance functions of the device gamut into the SVD-derived reflectance space; edges connect the (coarsely sampled) printer values to illustrated the volume.

The figure is analogous to the chromaticity coordinates plotted in Figure 6.3. As before, the grid lines interconnect tiles which are adjacent in their $RGB$ input specification. The figure may be created by identical means to the lattice employing CIEXYZ chromaticity coordinates, except that the reflectance vectors have been directly substituted for the tristimulus vectors and serve as the detector profiles by employing equations [2.16] and [2.29]. Orthogonality of the basis elements once again are useful as the mapping may take advantage of the matrix defined in [4.4].

As before, the overall geometry of the figure helpful in studying overall unifor-

Figure 6.11: SVD Reflectance Gamut (2-D projections)

Figure 6.12: Illumination versus Reflectance Coordinates (3D)

mity of density. A companion set of three scatter plots appear in Figure 6.11 which plot the distribution of points by choosing two of three axes. That is, they are the projections of Figure 6.10 onto the $x$, $y$ and $z$ coordinate axes.

### Illumination versus Reflectance Coordinates

A direct comparison of both methods appears in Figure 6.12. In 6.12(a) the uniform $4 \times 4 \times 4$ lattice of $[RGB]$ printer coordinates appears. In (b) and (c) the reflectance functions of each tile mapping onto $(X, Y, Z)$ chromaticity or $(\beta_1, \beta_2, \beta_3)$

reflectance coordinates, respectively. The increased uniformity of (c) in comparison to (a) indicates less "coupling" between colour coordinates. That is, reflectance coordinates provide a practical means of recording surface colour which make good use of their domain of representation. The bunching of lines in some regions indicates per-channel print non-linearity. This is easily corrected by overriding the factory-specified hardware lookup tables with gamma corrected or otherwise linearized values.

### 6.3.4 SVD Reflectance Matching – Mosaic

The mosaic provides a useful set of reflectance data for testing the methods of matching. Since this data is a coarse sampling of the gamut, it may be considered a self-test. The reuse is helpful because reflectance (coordinate) data sets are not yet common in most graphics laboratories. Commercial digital scanners do not provide full-resolution multi-dimensional imaging[9].

As an initial check, the corner tiles forming the convex hull of the gamut were reused. By expressing each of the eight primaries in reflectance coordinates and then back onto continuous functions, the original and approximating reflectances may be compared. These appear in Figure 6.13.

The comparisons are annotated with the chromaticity coordinates of both curves under source E. In addition, the CIE total colour difference $\Delta E^*_{uv}$ computed from the CIE 1976 ($L^*u^*v^*$) space give an indication of perceptual colour difference under this source. Under this scale the unit distance is just perceivable; differences in the range of five to ten are customarily considered a close match (Schwarz, Cowan and

[9]Landsat and other remote-sensing satellites provide multi-spectral scanning.

Figure 6.13: SVD Reflectance Matches of the Gamut Corners

Beatty, 1987). The software allows for additional measures, such as unweighted least-squares difference.

The last figure is the approximation to the ideal white, since the unprinted paper is considered uniformly 100% reflectance (a consequence of normalization). As predicted in connection with the SVD solid (Figure 6.9), this sample violates the reflection law [2.3]. Since the white sample lies (at the edge of) the printer's gamut this tile represents a point in which the SVD-derived approximating gamut is not strictly contained within the SVD-derived reflectance solid. This is the method's major liability.

Note in passing that the goodness of the white approximation gauged by the metric remains high, in part because the excursions are both balanced about $y = 1$ and of reduced amplitude through the mid-range. This is misleading: near-white paper stock will provide the most difficult matches, as their reflectances are most apt to lie outside the colour solid. (Print matching of the stock is deferred until the following section and MSSM-based methods).

### 6.3.5 SVD Basis Chromaticities

Although not a rigorous application of the CIE chart, the chromaticities of the reflectance functions themselves, $\beta_1(\lambda) \ldots \beta_4(\lambda)$ may be plotted. This conceptualization is suggested in §6.3.2. Other researchers of note have employed illumination-based CIE metrics in this fashion (Buchsbaum and Gottshalk, 1984). In Figure 6.14, the chromaticity associated with each reflectance element $1 \leq k \leq 8$ is plotted and annotated with numeral $k$.

Although the associated hues may lie outside the chromaticity diagram, their relative position indicates their contribution in forming reflective surfaces of varying

Figure 6.14: SVD Reflectance Element Chromaticities

colour. Because any non-black reflectance has a positive contribution of the non-negative basis element $\beta_1(\lambda)$, a near-white point is located near the centre of the chart. By linearity considerations, contributions of other elements perturb the point along the common line. Unlike the rules of additive composition, the motion may be toward or away the new source, whose contribution may have either sign[10].

Since the dimension of the deflationary space is fixed at three (the dimension of the printer's controls), the omission of the fourth basis vector weakens the approximation for surfaces having cyan (red/anti-red) components.

[10]In this sense "black light" becomes "negative light".

### 6.3.6 Conclusions

Global deflation based on SVD-derived orthogonal bases was applied to the Kodak SVC6500 printer. The goodness of fit is gauged both by using standard colorimetric tools and by visual estimation. SVD is a viable technique: it provided a variance minimization to the entire data set (Wesolowsky, 1976) which decreased with increasing dimension. The method of reflectance matching compares favourably to traditional methods of print matching (by illumination coordinates).

The method has one severe liability: its failure to approximate near-white reflectances (as are common with paper stock) with vectors representing physically possible surfaces. That is, points of approximation within the reflectance gamut may lie outside the reflectance solid; the containment of the former by the latter, as with the analogous containment of (additive) print gamuts within the chromaticity diagram, is lost. This shortcoming was studied in §4.2.7 and provides motivation for the MSSM-derived basis employed in the following section.

## 6.4 Print Matching – Global MSSM Method

The fifth chapter presented an algorithm [5.1.5] which derived SVD-like orthogonal basis elements $\beta_2 \ldots \beta_n$ in the presence of a constrained initial basis (eigen)vector $\bar{\beta}_1(\lambda) = c$. In theory, the basis modification introduces useful symmetry. In particular, the white point (conjugate to the black apex) now exists for any MSSM reflectance space of any dimension. In practice, this subsection describes empirical results in applying the MSSM (modified) method. After providing further motivation, the section closely parallels research and exposition of the previous section. As before, the conclusions provide impetus for the following section.

### 6.4.1 MSSM Printer Gamut

Establishing the grey axis as the principle reflectance element of the MSSM method (regardless of the profile of the function as derived by SVD) is strongly suggested by practical reasons. Commercial reprographics places a great emphasis on the *equivalent neutral density* or *END* of both colour and monochromatic print (Pobboravsky, 1966), (Pobboravsky and Pearson, 1972) media and photography (Evans, Hanson and Brewer, 1953). These methods methods provide for hue invariance, as higher-level cognative processes in the human visual system are highly responsive to hue shifts, especially across colour category boundaries (Boynton and Gordon, 1965), (Tansley and Boynton, 1977), (Koenderink, 1987), (Smallman and Boynton, 1990).

By placing the grey axis wholly within the (lowest dimensional) reflectance space, modeling of an "ideal grey" constant reflectance $\beta(\lambda) = c$ will be exact at all dimensions. This greatly improves the utility of the method when matching paper stock which shows little spectral bias. The "stock white" reflectance lies within the printer's gamut and is produced in the absence of colouring matter. Generally, four-colour CMYK printing produces a greatly improved grey tone by *undercolour removal,* that is, the substitution of the black "K" media with proportional reduction in the CMY inks. Because black colouring matter well approximate the ideal black reflectance function $\rho(\lambda) \approx 0$, the range of greys also shows strong spectral invariance. Thus, printer grey lies (nearly) within the MSSM reflectance solid of arbitrary dimension. Simply put, the gamut of the four-colour printer is well-contained within a MSSM-derived reflectance solid.

### 6.4.2 Principle Element Adjustment

The initial basis element $\tilde{\beta}_1(\lambda)$ previously derived by SVD from the three-colour Kodak printer shows little deviation (Figure 6.15) from the ideal element $\bar{\beta}_1(\lambda) = c$ with constant $c = n^{-\frac{1}{2}}$ to provide unit length. Given a printer having good gray-balance, it is not surprising that a suitable linear combination of the first three elements model a near-constant reflectance. What is surprising is that the first eigenvector $\tilde{\beta}_1(\lambda)$ *alone* approximates this function well. That is, the adjustment vector $\|\tilde{\beta}_1 - \bar{\beta}_1\|$ is small; the projected length $\tilde{\beta}_1 \cdot \bar{\beta}_1$ close to unity. This suggests that the dominant feature of a uniform gamut sampling is the surface "lightness" having little spectral bias. This is suggested by studies of Munsell chip reflectances (Cohen, 1964), (Parkkinen, Hallikainen and Jaaskelainen, 1989) and is confirmed by a contemporary analysis of the principal components of the reflectance functions of the naturally-occurring surfaces (Maloney, 1985a), (Maloney, 1986), taken from a well-known source (Krinov, 1947).

### 6.4.3 MSSM Basis Determination

The methods described in §5.1.5 were applied to provide an orthogonal set of discrete reflectance vectors. Weilandt deflation of the covariance matrix **S** by the constrained vector $\bar{\beta}_1(\lambda) = 1$ "broke" the subsequent iteration scheme, which found the dominant eigenvector using the Power method, presented here as pseudocode. (See also the comments in the C-language source code of Appendix A).

```
v <- [ . . . ]      (initialize to arbitrary value)
   repeat
   v' <- Sv         (matrix premultiplication)
```

```
v' <- v'/|v'|   (normalization)
v <- v'
while (v' <> v). (until converges)
```

The while loop completes when the system converges. The stopping condition $\mu\mathbf{v} = \mathbf{S}\mathbf{v}$ are the defining criterion for a (right) eigenvector $\mathbf{v}$ having associated eigenvalue $\mu$. For completeness sake, a second loop exit is provided should convergence fail by the hundredth step. This code was tested accidentally: the "arbitrary" initialization vector used in previous SVD application set all components: $\mathbf{v} = (1, \ldots, 1)$. Since this vector is the modified basis reflectance $\bar{b}_1$, the preadjustment to the data matrix $\mathbf{S}$ removed this grey axis, placing the initialized $\mathbf{v}$ within the null space of the rank reduced $\mathbf{S}$. This not only exercised the guard section of the code but verified the stability of the technique: numerical imprecision in the matrix multiplication would otherwise perturb successive approximations $\mathbf{v}'$ out of the null space and an eigenvector might become located (Cook and Weisberg, 1982). (The eigenvector would be correct in value, but the algorithm then incorrect in operation.) This stability may be explained in part by the use of double-precision calculations on intermediate calculations (Rice, 1983) together with a dot-product operation which tallies positive and negative element products independently, so that only one (and not $k$) summations of quantities of unlike sign take place in calculating each product element, thus maintaining precision within the floating point mantissa (Morris, 1983). In passing, this was the only time the iterative method failed to converge, as empirical data never presented pathological starting conditions in the form of two dominant eigenvectors having nearly identical eigenvalue.

Plots of the first six MSSM reflectance vectors appear in Figure 6.15(a) and (b). Vectors other than the first are annotated with their matching eigenvalue.

Figure 6.15: MSSM Reflectance Vectors

By comparison to the SVD method, the values associated with each reflectance vector is larger owing to the presence of the constraint. When used with small systems ($m = 3$) some degradation in representational loss can be expected. This is summarized in Table 6.1.

### 6.4.4 Representational Loss

A comparative listing of representational accuracy as a function of elements retained may be tabulated for both the SVD and MSSM methods. The entries are the (natural) logarithms of the eigenvalues associated with each principle component. This maps the stopping condition $\mu = 1$ to zero as $\log 1 = 0$ regardless of base, which would scale each entry. Moreover, recording instead the squared-variance also scales (doubles) each entry. In each case the alternate stopping criterion is also preserved: a large ratio between successive two entries becomes a large difference.

No value is associated with the first component under the MSSM method. Instead, an additional sixty-fourth component is produced after decomposing the preadjusted system. Its logarithm was $-32.09$ ($= \ln\ 1.1538 \times 10^{-14}$), indicating the algorithm's success in detecting the null space lost through imposition of the first eigenvector. Although sampled at a resolution $n > 64$, sixty-four components provide sufficient degrees of freedom to represent any reflectance in the $4 \times 4 \times 4$ gamut exactly.

Note that the values agree (coincidentally, they are identical) at the fifth entry. The sixth entry is then near/below unity and additionally shows a large step. Thus, the alternate formulation quickly wins back any "ground" lost in representational accuracy. Since the SVD method is provably the lower bound on variance minimization, the "slack" introduced by constraining the solution presents little loss

| Element Number | SVD Method | MSSM Method |
|---|---|---|
| 1 | 3.5812 | —— |
| 2 | 2.3275 | 2.4985 |
| 3 | 1.7310 | 1.7383 |
| 4 | 0.8252 | 1.0079 |
| 5 | 0.5669 | 0.5669 |
| 6 | -0.0005 | 0.0345 |
| 7 | -0.1072 | -0.0360 |
| 8 | -0.1859 | -0.1153 |
| ... | ... | ... |
| 64 | -6.0560 | -32.093 |

Table 6.1: Paper Variance Minimization: SVD versus MSSM

in accuracy. The higher-dimensional reflectance spaces have nearly identical span (they exclude a nearly common null space), though the MSSM-derived space retains the grey axis across its full range of dimension. The higher dimensional solids thus nearly match to within rotation and reflection. For the space $m = 3$ any loss in representational accuracy is make up for by the space's ability represent the white point (grey axis) properly.

It should be emphasized that the MSSM method meets both defining criteria for useful deflation outlined at the chapter's start. Though no longer strictly eigenvectors, the basis elements produced are orthogonal and presented in order of decreasing goodness of fit, forming a nested sequence of reflectance spaces of increasing representational power. Although the modified MSSM method ranks only the second element onward, the strong similarity of initial elements $\tilde{\beta}_1(\lambda)$ to $\bar{\beta}_1(\lambda)$

Figure 6.16: MSSM Reflectance Solid

suggest that that latter properly belongs in primary position.

### 6.4.5 The MSSM Reflectance Solid

The MSSM reflectance solid for the Kodak printer, illuminated by CIE source E, is presented in Figure 6.16 (compare Figure 6.9). As before, the figure is created as a screen capture and colour hardcopy of the SGI/Iris workstation, for which (alone) the colours are correct in a self-referential manner.

#### Rendering the MSSM Solid

The solid is illuminated by CIE source E. The linear combinations of surface reflectances are thereby "illumined"; the method of conversion parallels the CRT model described in §3.1.2. In contrast to [3.4], the pixel vector of the right-hand side of [6.9] presents illumination coordinates[11] which are tristimulus values. For the MSSM-derived reflectance space, the transformations are:

[11]This matrix-based CRT model is extended further in §8.2.1 and following.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 25.93 & -13.16 & 4.55 \\ 46.21 & -40.74 & -5.47 \\ -8.91 & 16.28 & -7.37 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad [6.9]$$

and

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} .0577 & -.0034 & .0382 \\ .0577 & -.0223 & .0522 \\ .0577 & -.0452 & -.0665 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \quad [6.10]$$

Note the constant initial column in the matrix in [6.9]. This indicates that the equipotential surface represented by the reflectance coordinate column vector $[c, 0, \ldots, 0]^{\mathrm{T}}$ produces tristimulus values $x = y = z = c$ under source E and hence the white chromaticity coordinates $x' = y' = \frac{1}{3}$. Whether the CRT provides a colour which truly appears grey is a question of second-order considerations such as gamma correction and linearity adjustments in each video channel's final lookup table (LUT), which on this hardware is distinct from the colour lookup tables (CLUT).

#### MSSM Reflectance Coordinates

As in Figure 6.5, a scatter plot may be used to study the domain of the coordinates. Here the MSSM printer gamut is projected orthographically onto the $x$, $y$ and $z$ planes lying in 3-D reflectance space. This appears in Figure 6.17.

### 6.4.6 Features of the MSSM Solid

The new solid is no longer "snubbed", as described previously. The grey axis and $x$ axes are identical, joining at the white apex. This vertex is *degenerate* (more than

Figure 6.17: MSSM Reflectance Gamut (2-D projections)

the $m$ requisite $m$-hyperplanes define it) and it forms a conjugate vertex with the antipodal black apex, by reason of central symmetry. In fact, it is a point common to *all* $n$ zone boundaries (with $n$ the number number of wavelength samples and $n > m$), though at each zone's upper limit. In general, any vertex on the surface of the MSSM solid pairs with a conjugate vertex having an identical set of zone intersections in which the "parity" of the upper and lower surface of matching zones are reversed.

As before, points along the surface describe reflectance functions bounded by the condition $0 \leq \rho(\lambda) \leq 1$ in which the conditions of equality are met at exactly $i$ points for a cell of dimension $m - i$. The "belt" along the solid defines the set of vertices having "mixed parity" in terms of upper and lower zone membership. That is, this circumference is the locus of all reflectance functions which are simultaneously bounded by both the upper and lower limits of physical reflection. The belt's lack of $C^1$ continuity was described in §4.2.4. Further exploration of the belt and its associated reflectance is taken up in the next chapter.

### 6.4.7 MSSM Reflectance Matching – Mosaic

The mosaic once again provides a self-test of the methods. The reflectance vectors of the mosaic tiles corresponding to the eight corners of the gamut are approximated using the MSSM basis. Plots appears in Figure 6.18. As before, they are annotated with the CIE total colour difference $\Delta E^*_{uv}$.

Note that all approximations satisfy the reflection law [2.3] and in particular the white reflectance is represented exactly. That is, the gamut lies within the MSSM solid. This overcomes the liability of the SVD method suggested by first-principles of regression. However, the total colour difference has grown substantially for sample having substantial reflectivity at long wavelengths (red, magenta). This is anticipated given the chromaticities of the first three MSSM reflectance elements appearing in Figure 6.21.

### 6.4.8 MSSM Reflectance Matching – Paper Stock

Paper stock was characterized using spectroradiometric methods identical to the procedure for gamut sampling. This included a uniform lighting geometry in both cases: source forty-five degrees from vertical, aperture vertical (normal) to the sample's surface. Their reflectance functions together with an approximating reflectance appears in Figure 6.19.

The plots are analogous to those in Figure 6.18 save for the substitution of the original surfaces under characterization. (In some bound copies of this thesis the spectra of Figure 6.19 are printed on the paper stock from which they were derived, forming a useful self-referential artifact).

As the reflectances presented are not necessarily reproducible on the printer,

Figure 6.18: MSSM Reflectance Matches of the Gamut Corners

Figure 6.19: MSSM Reflectance Matches of Paper Stock

| Stock | Elements | | | |
|---|---|---|---|---|
| Colour | 1 | 2 | 3 | 4 |
| grey | 25.3 | 2.42 | *1.86* | 2.62 |
| canary | 73.0 | 69.7 | *11.1* | 18.8 |
| pink | 22.2 | 20.0 | *16.0* | 6.12 |
| palecyan | 52.0 | 14.2 | *13.1* | 10.6 |
| ltorange | 61.5 | 43.6 | *16.8* | 2.34 |
| lime | 83.8 | 81.0 | *19.0* | 23.5 |
| aqua | 84.8 | 33.4 | *20.3* | 20.0 |
| orange | 92.2 | 29.7 | *35.8* | 8.19 |

Table 6.2: Paper Stock $\Delta E^*_{uv}$ by Reflectance Space Dimension

the approximations deviate more than the self-test. Problem specifics allow the retention of only three vectors. An increase in representational accuracy may be tabulated as a function of number of basis elements. A summary of the total colour difference $\Delta E^*_{uv}$ for each stock under an imaginary source E appear in Table 6.2. The $n$th column indicates goodness-of-fit when the approximating space retains the first $n$ reflectance vectors.

As seen, the method provides an excellent match in all cases with the introduction of the fourth element. This suggests that the non-linear behaviour of the three-dimensional $RGB \rightarrow CMY$ printer transfer function can be modeled with the addition of just one parameter. Unfortunately, a model having only three degrees of freedom is sought to make a unique inverse mapping possible.

**MSSM Matching – Best/Worst**

The successive reflectance approximations for the stock having best fit (grey) and worst fit (orange) are presented in detail in Figure 6.20. In each instance the first (solid) curve is the original reflectance, the second through fifth curves giving approximations using the MSSM-derived orthogonal reflectance spaces having from one to four dimensions.

Note that the stock sample (grey) most resembling the ideal white is very well-behaved. It converges quickly and exhibits no $y > 1$ excursions. Only the orange sample violates the physical condition, albeit at the opposite $y < 0$ boundary. This suggests a loss in the degrees of freedom for specification of spectral profile in the portion of the spectrum appearing between the 444nm and 450nm graticules. A similar deviation is seen in the previous Figure 6.13(d) for the cyan sample, whose peak reflectivity also falls within this range.

### 6.4.9 MSSM Basis Chromaticities

As before, the CIE chromaticity diagram in Figure 6.21 records the "hue" of each basis element. The difficulty in controlling the cyan sample (or its complementary red-orange) arises because there is reduced control along this axis. (Note that the principal components derived may be reversed about the white point in comparison to the Figure 6.14 for reasons described in §4.2.2.)

### 6.4.10 Chromaticity Migration

Hue migrations of the paper samples may be plotted as vectors on the chromaticity diagram (this is a more appropriate use of the chart). These appear in Figure 6.22.

Figure 6.20: MSSM Reflectance Matches of Paper Stock (best/worst)

Figure 6.21: MSSM Reflectance Element Chromaticities

The chromaticity migration for each sample is plotted as a vector whose head indicates the chromaticity of the approximation and whose length indicates migration from the exact value. The approximation is conducted on reflectance spaces ranging from first through fourth dimension; the vector tail (and annotation) remaining fixed. The CIE 1976 ($L^*u^*v^*$) space, used previously to compute colour differences, is the space of choice. Since the transformation from $(x', y')$ onto $(u', v')$ space is affine, any linear features within the tristimulus (and reflectance) space are preserved. In Figure 6.22(a), plotting all samples under a single basis vector produces a class of grey surfaces which vary only by overall reflectivity (are metameric with regard to colour). All arrowheads converge at the white point $x' = y' = \frac{1}{3}$ which defines CIE source E. In (b) the added degree of freedom introduced by the second element produces approximations whose locus of chromaticities plot as a line. In (c) three elements are used; the approximate chromaticities are the reflectance matches studied previously. Note that the vector corresponding to the orange sample has

Figure 6.22: MSSM Paper Chromaticity Migration

lengthened, indicating that an additional reflectance element worsens this particular sample under the chosen source of illumination. This result is both important and counter-intuitive. By happenstance, the deviation of the orange match splits the difference of the desired reflectance, producing an unusually low total colour difference (see in particular Figure 6.20). Finally the inclusion of the fourth vector drives the system nearly to convergence.

### 6.4.11 Conclusions

Global deflation based on MSSM-derived colour gamuts was applied to the Kodak SVC6500 printer. The methods satisfied the defining criteria for useful deflation. Besides theoretical advantages of "perfect" white and symmetry, a larger percentage of the printer's gamut now lay within (the subspace defined by) the reflectance solid.

In addition, any initial loss in representational accuracy is quickly offset, making the MSSM-derived basis the method of choice for general modeling of reflectance functions with a small number of parameters.

The only liability with the method of reflectance matching is in loss of representational accuracy at very low dimensions. Although the MSSM quickly matches the accuracy of the SVD method with increasing dimension, neither can adequately model the printer's transfer function using a global three-dimensional model. This suggests a means of local analysis, described in the following section.

## 6.5 Local MSSM Deflation

Methods of local deflation are now introduced to provide additional representational accuracy. These preserve the desirable properties of MSSM-based deflation while offering tighter fits of the reflectance space in regions where the printer gamut is not well-characterized. Local bases are employed to the specific problem at hand: the matching of surfaces by reflectance function. Exposition defines a methodology for its solution and cannot parallel the previous two (e.g., no canonical reflectance solid exists for the printer gamut).

This approach is the method of choice for two reasons. First, the SVD-based method established a lower bound having marginal representational accuracy at three dimensions. Improvement thus requires local methods. Second, application of the MSSM method of basis derivation is desirable. Any further constraining of basis elements introduces spectral bias an a basis element by reason of orthogonality to $\bar{\beta}(\lambda)_1$. Any choice can be literally "cast in a bad light" by an adversary (Rawlins, 1992) choosing a source of illumination based on an *a priori* definition of the basis element. Thus, local deflation is required.

A method of refinement to the reflectance-matching experiment is desired. Because the dimension is fixed at three, additional control over the behaviour of the print function is made possible by subdividing the print space and characterizing each separately. This can quickly lead to large spaces, even a binary subdivision on each axis provides $2^n$ regions.

Local deflation introduces spatial dependency to the basis elements for the printer gamut, thus allowing retention of the three-parameter model. Multiple basis sets raises new concerns. First, the method and degree of the partitioning of a reflectance space must be established. Second, means of membership testing is required to identify the best partition when presented a reflectance function for surface matching. Both are treated below.

### 6.5.1 Octant Creation and Continuity

Though smaller partitionings are possible, a simple and symmetric approach bisects each of the three axes which define the printer gamut data. This forms a $2 \times 2 \times 2$ Cartesian partitioning having $2^m$ congruent parts, where $m = 3$ (the dimension of the reflectance space). Though the number of (hyper) octants grows rapidly in higher dimensions, eight is not severe; at higher dimensions the global methods may be substituted.

A MSSM reflectance space (solid) is derived for the subset of mosaic tiles falling within each octant. This breakdown is useful, suggesting that finer partitioning is unnecessary. First, each of the conventional primary and secondary colours[12]. are represented by a unique subcube. Thus, the physical interactions defined by the power set on the $\{cyan, magenta, yellow\}$ printing primaries are in one-to-one

[12]These form the set $\{black, red, green, blue, cyan, magenta, yellow, white\}$.

correspondence with each subcube. That is, the interactions and in particular any non-linearities of any given regime of ink overprinting, such as "cyan+magenta" are restricted to a single subcube, here "blue" (minus-yellow).

This power set defines the *reproduction primaries* (Pobboravsky, 1966) which customarily forms the test strip printed as fiducials on the margin of stock purposes of both colour correction and registration. The benefits of models based upon this partitioning (Neugebauer, 1937) previously found wide-spread application in pre-digital graphic arts. In addition, regions of *overprint,* customarily considered excessive at 280% (that is, when 2.8 of 3 possible full ink units are imaged) belong to the "black" subcube. Thus, the modeling of regions of high print non-linearity are restricted to this subcube. Though not applicable with the printers under study, undercolour removal (Yule and Claper, 1960) can also remain largely local to this subcube; extended methods harmonize well with this systematic approach.

Related colour correction work (Starkweather, 1988) samples a 3-dimensional printer gamut merely along its faces or edges. This reduces the number of parameters from cubic to quadratic in proportion to the axis partitions. This provides excellent correction in regions of non-linearity near the corners, while allowing for simple linear interpolation to "fill-in" the body of the cube (where the print function is much better behaved). However, the method is *ad hoc* in that mid-range print balance (non-linearity correction) is left untreated. Worse, there is no assurance of even $C^0$ continuity between the mappings defined at adjacent estimates. Such discontinuities are highly objectionable and have plagued commercial hardware products (Keightley, 1987) when round-offer error appears in integer-based linear interpolation (this is explained in detail in Appendix H and also treated in (Paeth, 1989b)). Even tone gradations which are $C^0$ continuous but $C^1$ discontinuous produce objectionable Mach bands (Ratliff, 1965) because the human visual

system is highly attuned to changes in surface brightness.

In contrast, the presence of the uniform reflectance vector $\bar{\beta}(\lambda) = c$ is a boon in providing a well-defined grey axis running as the body diagonal across the printer gamut. The point of eight-fold join at the centre of the $2 \times 2 \times 2$ lattice defines the equipotential 50% grey, which is exactly representable at the vertex common to all eight subcubes. Deviations away form this point of symmetry (into each proper volume of each subcube) thus do not exhibit any discontinuity, which would be quite evident at these midrange tonalities.

### 6.5.2 Local MSSM Bases

The $4 \times 4 \times 4$ Kodak SVC6500 print mosaic was too sparse following this three-fold subdivision: only eight samples existed in each subcube. A $8 \times 8 \times 8$ gamut was created as a mosaic of eight sheets each having eight tiles[13]. Both the SVD and MSSM bases were derived on the entire data set of 512 tiles. This is a first overdetermined system: the number of tiles exceeds wavelength dimension ($n = 300$ on the domain 400nm to 700nm in 1nm steps). A comparative chart of the first three reflectance vectors produced by each method appears in Figure 6.23.

The strong similarity of the two bases further confirms the utility of the MSSM method – the representational loss in modeling the printer is largely a consequence of restricting the problem space to $m = 3$. For the remaining experimentation, merely MSSM-derived bases are used. Note in passing that second reflectance vectors differ by sign reversal. The "normalization" technique of negating any reflectance vector by making its first element positive fails to present the same "handedness" for similar vectors created from dissimilar data sets. The normaliza-

[13]This was the largest characterization of the SVD6500.

Figure 6.23: SVD vs MSSM Basis on Global $8 \times 8 \times 8$ Gamut

tions which conclude (Parkkinen, Hallikainen and Jaaskelainen, 1989) are similar and likewise fail. A potential remedy places each reflectance function's maximum extrema on the positive side of the $y$ axis.

The first three basis elements derived under the MSSM method are presented for each subcube in Figure 6.24. Each was derived sixty-four mosaic tiles using procedures identical to the previous mosaic.

### 6.5.3 Local MSSM Reflectance Solids

The related MSSM reflectance solids are plotted in Figure 6.25. Because of their self-referential nature, the "black" ("white") solids are necessarily dark (light). Each solid lies within a distinct reflectance space, though the grey axis is common to all. Note the similarity of the white local solid (bottom right) to its globally derived MSSM counterpart appearing in Figure 6.16. Two other solids are of interest (placement corresponds to their defining basis elements in Figure 6.24). The local green solid (second row, first column) resembles the bi-cone of Figure 5.5. The cyan solid (adjacent) exhibits the maximum girth of any reflectance solid studied. Its sharp dihedral break is a consequence of the end-to-end wavelength discontinuity of its second and third defining reflectance elements.

End-on views of the white apex aid in seeing which hue circuits are the least or most out of round. Three extreme examples appear in Figure 6.26. Each has been scaled 1.5×; as before all images use a common registration, inviting visual comparison of girth extend in reference to the box edge.

Note that the examples are the printing secondaries. Out of roundness need not indicate poor goodness of fit of the local basis. Should all printing primaries be well-represented in the local basis while utilizing merely the first two basis elements,

Figure 6.24: Local MSSM Reflectance Matches of the Gamut Corners

Figure 6.25: Local MSSM Reflectance Solids (8)

Figure 6.26: Red, Green and Blue Local MSSM Solids (Apex View)

then the profile of the third function (and its continuity in particular) becomes arbitrary, as the dominance of its axis vanished. This suggests a differential axis scaling in proportion to associated singular value. Orthonormality is lost but lines are preserved. The scaled reflectance space then supports a Euclidean distance metric or *reflectance measure* between reflectance coordinates useful in estimating similarity of reflectance functions. This is a direct extension of the affine transformation [3.28] which defines the CIELUV uniform colour space (Figure 3.3) and associated metric, the CIE total colour difference $\Delta E^*_{uv}$.

### 6.5.4 Print Matching Overview

The procedure for reflectance print matching is briefly recapitulated in this new context. First, a given reflectance function is mapped onto the reflectance coordinates of each octant. The octant which best approximates the reflectance is chosen as representative. (This choice is justified below.) An analytical method allowing substantial precomputation is derived to assist the search. This is important because reflectance mapping is a prime candidate for accurate digital imaging, and the number of pixels of a $2-D$ image grows quadratically with edge size, making the per-pixel computation time the dominant expense. Once the basis is identified

Figure 6.27: Lattice Plots of Best Approximating Bases

the method proceeds as with the global MSSM.

Print matching requires a source of originals presented in reflectance coordinates. Because reflectance-coordinate images are generally unavailable, advantage once again taken of the gamut reflectances to provide input stock.

### 6.5.5 Choice of Best Basis

A simplifying assumption is that the octant in which an arbitrary reflectance function has its closest fit is also the octant will ultimately define the reflectance to printer coordinate mapping. That is, the chosen inverse transfer function will call for a set of printer coordinates lying within the octant (portion of the printer gamut) corresponding to the reflectance space used. The assumption may be validated by scatter plotting the $8 \times 8 \times 8$ mosaic in 3-space. In Figure 6.27(a) a lattice is presented in which both the colour and location of each point is determined by octant in which each sample's print coordinates lie.

In (b) the print samples best represented when approximated under the "magenta" reflectance basis are plotted. The spatial location is again defined by the coordinates of the input sample; all points are plotted in magenta. As is clear, the

percentage of *rogue* points violating the assumption are small. Moreover, closer analysis of these points demonstrates that they are very well-approximated by a number of local bases: their migration into a neighboring local basis is of little concern. As can be seen, the assumption is reasonable. What remains is a method to identify the local basis of best fit for a given reflectance function.

### 6.5.6 Location of Best Basis

The method of local identification is derived from the global method, with an common subterms removed. That is, given a spectrum and a basis, a deflation onto three reflectance coordinates is created. These define an approximate reflectance function $\tilde{\mathbf{r}}$ lying within the reflectance space, whose least-squares difference with the given reflectance $\mathbf{r}$ may be computed by applying the scalar inner product (with cost proportional to vector length).

The linear algebra which underlies these operations is as follows. The projection (deflation) of $\mathbf{r}$ onto $\mathbf{B}$ is straightforward: the reflectance coordinates are $\tilde{\mathbf{x}} = B^{\mathrm{T}}\,\mathbf{r}$ (equation [4.5]) where $\mathbf{B}$ consists of orthogonal columns representing the basis set. The right-hand side dots the column vector $\mathbf{r}$ successively with each column in $\mathbf{B}$ to create each reflectance component in the reflectance coordinates. These projected lengths may be reexpressed as the form $\tilde{\mathbf{r}}$ conforming to $\mathbf{r}$ by reusing the basis $\mathbf{B}$, that is, $\tilde{\mathbf{r}} = \mathbf{B}\tilde{\mathbf{x}} = \mathbf{B}\mathbf{B}^{\mathrm{T}}\,\mathbf{r}$ (by [4.12]).

A means of simplying the operation for reuse is desired. Precomputation of $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ would be a mistake: a large symmetric matrix is then created. But $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}$. That is, the basis columns are orthonormal; the operation forms the identity in the dimension of the subspace. This identity is used in the formal derivation presented below.

### 6.5.7 Algebra for Basis Identification

Given: $\mathbf{B}$ a rectangular basis of orthonormal columns and $\mathbf{r}$ an arbitrary reflectance vector. Compute: the absolute error value $e$ as $\|\mathbf{r}-\tilde{\mathbf{r}}\|$ where $\tilde{\mathbf{r}}$ is the least-squares approximation of $\mathbf{r}$ on basis $\mathbf{B}$. This may be defined as the scalar-valued function $\Delta(\mathbf{B},\mathbf{r})$. Thus:

$$\begin{aligned}\|\mathbf{r}-\tilde{\mathbf{r}}\| &= [(\mathbf{r}-\tilde{\mathbf{r}})^{\mathrm{T}}(\mathbf{r}-\tilde{\mathbf{r}})]^{\frac{1}{2}} \\ &= [(\mathbf{r}^{\mathrm{T}}-\tilde{\mathbf{r}}^{\mathrm{T}})(\mathbf{r}-\tilde{\mathbf{r}})]^{\frac{1}{2}} \\ &= \mathbf{r}^{\mathrm{T}}\mathbf{r}-\tilde{\mathbf{r}}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\tilde{\mathbf{r}}+\tilde{\mathbf{r}}^{\mathrm{T}}\tilde{\mathbf{r}}. \qquad [6.11]\end{aligned}$$

Substituting $\tilde{\mathbf{r}}=\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}$ gives

$$\begin{aligned}\|\mathbf{r}-\tilde{\mathbf{r}}\| &= \|\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}-\mathbf{r}\| \\ &= [(\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r})^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r})-\mathbf{r}^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r})-(\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r})^{\mathrm{T}}\mathbf{r}+\mathbf{r}^{\mathrm{T}}\mathbf{r}]^{\frac{1}{2}}, \qquad [6.12]\end{aligned}$$

Since $(\mathbf{A}\mathbf{B})^{\mathrm{T}}=\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$, with $(\mathbf{B}\mathbf{B}^{\mathrm{T}})^{\mathrm{T}}=\mathbf{B}\mathbf{B}^{\mathrm{T}}$, expanding gives

$$= [\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}+\mathbf{r}^{\mathrm{T}}\mathbf{r}]^{\frac{1}{2}}$$

but $\mathbf{B}^{\mathrm{T}}\mathbf{B}=\mathbf{I}$, (though $\mathbf{B}\mathbf{B}^{\mathrm{T}}$ is not!), yielding

$$= [\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}+\mathbf{r}^{\mathrm{T}}\mathbf{r}]^{\frac{1}{2}}$$

with the first and second terms cancelling, giving at last

$$\Delta(\mathbf{B},\mathbf{r}) = [\mathbf{r}^{\mathrm{T}}\mathbf{r}-\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}]^{\frac{1}{2}}. \qquad [6.13]$$

In which all terms are scalars. This completes the derivation. ∎

### 6.5.8 Algorithmic Implementation

Least-squares introduces inequalities of squared variance, such as $\sigma_a^2 < \sigma_b^2$. Since the function $\sqrt{x}$ is strictly monotonic, the square-root may be removed when merely the inequality (and not its absolute value) are tested, saving cost. Next, the computation is sped by the absence of any the basis dependence $\mathbf{B}$ in the initial $\mathbf{r}^{\mathrm{T}}\mathbf{r}$ term. This scalar is the vector length of the spectral vector and is precomputed outside the iteration loop. The remaining $\mathbf{r}^{\mathrm{T}}\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{r}$ term may be computed in two steps. Rewriting the form as $(\mathbf{B}^{\mathrm{T}}\mathbf{r})^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}\mathbf{r})$ reveals that it is a dot product operating on a common subexpression. The inner loop evaluation is therefore in three parts. First, the subspace projection $\mathbf{B}^{\mathrm{T}}\mathbf{r}$ is computed, forming a small row vector. Next, the squared length of this vector is evaluated. Finally, a scalar difference is computed. Note that this subtraction cannot give a negative result as the underlying algebra implicit in $\|\mathbf{r} - \tilde{\mathbf{r}}\|$ is a positive length. The algorithm is presented below.

```
nearbasis(v: vector, bases: set of matrices)
   BEGIN
   vdot, pdot, mindist: REAL; m, : MATRIX;
   vdot := vector_length(v);
   mindist := 1.0e30;
   FOREACH m FROM bases
      BEGIN
      curdist := vdot - vector_length(mat_mult(mattransp(M),v))
      IF (curdist < mindist)
         BEGIN
         mindist := curdist;    --- 'mindist' is the squared error,
         close := m;            --- 'close' is the basis of best fit
         END
      END
   END
```

### 6.5.9 Discussion of Method

Intuitively, the operation $\mathbf{B}^{\mathrm{T}}\mathbf{r}$ rotates $\mathbf{r}$ as $\mathbf{B}$ is orthonormal. Since $\mathbf{B}$ is not square, the length of the rotated length is diminished as higher order terms are deflated. When the rotation places only a small contribution of the elided terms in the null space of $\mathbf{B}$ the length is essentially unchanged and the difference approaches zero, an indication of close fit.

This algorithm is asymptotically fast for finding the nearest basis. This is clear because it does the minimal work necessary: the components of each basis element in the test set are examined only once in forming the matrix multiply with the test spectra; the remaining steps are conducted on a vector of constant length. However, given a systematically created set of bases, a more direct classification may be possible after directly examining the input spectra without performing computations on all bases. This would require further study of rogue points on the lattice.

### 6.5.10 Local MSSM Reflectance Matching – Mosaic

The mosaic reflectance samples corresponding to the corners of the gamut were again matched using the technique. The comparison appears in Table 6.3.

It should be noted that the global SVD was from a previous (8× sparser) data set; Figure 6.23 strongly suggests that the set of global vectors are possibly equivalent. Comparing across data sets, the global MSSM method loses substantial ground to the SVD method for those samples containing substantial (minus) red[14].

[14]It is conjectured that the extended 370nm-740nm sampling domain is a liability, see §8.3.1 (last subsection).

| Gamut Sample | Global SVD | Global MSSM | Local MSSM |
|---|---|---|---|
| black | 4.46 | 12.99 | 1.49 |
| blue | 12.06 | 9.97 | 7.75 |
| green | 26.00 | 17.34 | 11.79 |
| cyan | 6.67 | 53.61 | 6.81 |
| red | 23.26 | 86.32 | 13.56 |
| magenta | 17.3 | 75.24 | 24.87 |
| yellow | 7.56 | 51.96 | 10.01 |
| white | 5.19 | 0 | 0 |
| (mean) | 12.81 | 38.43 | 9.54 |

Table 6.3: Gamut Variance Minimization: Comparisons

The local MSSM method has the best average of any technique. More importantly, the variance is reduced as well.

The total colour difference of the cyan (minus-red) sample is excellent and near idential for both the global and local MSSM methods. Its reflectance matches (on the reduced doman 400nm-700nm) are plotted using a basis of both two and three elements. As seen in Figure 6.28 the profile of the approximation is high.

A comparison may be made between the global and local methods by overlaying two histograms, each depicting least-squared variance. The sixty-two bins each contain an interval of $10^{-4}$ variance, their height is directly proportional to population. The histograms appearing in Figure 6.29 account for 95% of all data. The extreme bin of the global method (not shown) lies $3.48\times$ further than the last bin of the local method. The rogue entries, which account for 6.8% of the data points are not troublesome – they typically lie within the region of overlap where

Figure 6.28: Local MSSM Reflectance Match of Cyan

Figure 6.29: Histogram of Match Variance: Local vs Global MSSM

Figure 6.30: Artifacts: Original and MSSM Reflectance Match

classification does not matter.

### 6.5.11 Print Artifacts

The methods described were used to map reflectance coordinates for two selected samples onto printer input coordinates by using (inverse) tri-linear interpolation of the printer transfer function. The original paper and matching artifacts appear in Figure 6.30.

The matches were created using RGB printer input values of (180,186,186) for "grey" and (130,224,163) for "palecyan". Note that the algorithm has reduced the R component of the printer coordinates for grey in order to reduce an inherent pink bias in the appearance of the surface. This suggest on-going study in the behaviour of the printer *per se,* though this lies outside the scope of the thesis.

### 6.5.12 Conclusions

Local SVD provides a straightforward means to address the problems of colour correction in a reflective context. Specifically, variance has been greatly reduced while maintaining both the uniform grey axis plus general continuity of approximating reflectance functions. The method produces a viable low-dimensional model used to characterize the printer's transfer function.

## 6.6 Final Conclusions

Deflation has been successfully applied to the task of print matching. A methodology has been created by which reflectance bases may be used in procedural methods of printing. The technique is original in that matching of reflectances takes place without reference to a source of illumination. The methods are precise (demonstrated), though restricting the solution space to three-dimensions engenders a number of problem-specific techniques necessary to achieve the desired ends. With these restrictions removed it is evident that modified reflectance spaces of low (four to five) dimensions form highly accurate models whose total variance rivals the lower bounds established by standard regression methods. The MSSM modification introduces desirable advantages. These include symmetry of the reflectance solid, containment of the print gamut within this solid for near-white values and the consistent preservation of the grey axis of representation.

# Chapter 7

# Inflation

Inflation describes the linear transformation of reflectance coordinates onto a reflectance space of increased dimension. It is therefore a near complement to deflation: it defines a mapping which subsequent deflation inverts exactly, though the converse does not hold. This is analogous to the integration of polynomials, an operation which increments the degree of the integrand while introducing an unspecified constant of integration. Inflation, though invertible by deflation, is likewise not unique: information is introduced. Useful inflation locates a best axis of dimensional increase based on general properties of reflectance spaces and upon the nature of the problem at hand.

In practice the basis elements of the new space may not be uniquely determined. Consequently, the analytic methods presented in §7.2 provide a deterministic but necessarily partial solution to general inflation. Interactive exploration of the reflectance solids and their projections is conducted in §7.3. A custom solid modeler is described which aids in the visualization of solids. The modeler also helps develop intuition regarding reflectance spaces, leading to the creation of companion subwin-

dows used in the modeler. In §7.4 these are applied to study the solid's reflectance belt: a geometric feature which encodes much of the information of reflectance solids in a compact, lower-dimensional model. Analytic, then interactive methods allow for a rigorous determination of a candidate set of reflectance bases, with final choice made interactively, thus accommodating most problems. Finally, the interactive tools support modeling in higher-dimensional spaces, explored briefly in §7.5. The chapter concludes with a brief summary.

## 7.1 Overview

Inflation forms a complementary operation to deflation: given an $d$-dimensional reflectance space spanned by basis $\mathbf{B}^{(\mathrm{d})}$, the inflation $\mathbf{B}^{(\mathrm{d})} \xrightarrow{\mathrm{I}} \mathbf{B}^{(\mathrm{i})}$ finds the coordinate transformation onto $i$-dimensional space spanned by $\mathbf{B}^{(\mathrm{n})}$. Since both reflectance spaces are linear, the mapping takes the form of a linear transformation $\mathbf{T}$. In its basic setting:

$$\mathbf{B}_{n\times i} = \mathbf{B}_{n\times d}\mathbf{T}_{d\times i}, \qquad [7.1]$$

determines the new bases by post-multiplication of $\mathbf{B}^{(\mathrm{n})}$ by $\mathbf{T}$. This transformation matrix will be orthogonal *iff* both $\mathbf{B}^{(\mathrm{d})}$ and $\mathbf{B}^{(\mathrm{i})}$ are orthogonal.

In most cases the bases are not fully defined. The central concern of inflation then lies in establishing a useful linear transformation $\mathbf{T}$ given one partially defined and one fully defined basis. There are two cases: an unknown starting basis $\mathbf{B}^{(\mathrm{d})}$ and unknown ending basis $\mathbf{B}^{(\mathrm{i})}$. These are treated in turn.

### 7.1.1 Final space unknown (Case 1)

In the first case the larger space $\mathbf{B}^{(\mathrm{i})}$ is not fully defined. The first $d$ columns may be initialized to the values appearing in $\mathbf{B}^{(\mathrm{d})}$. Inflation then finds the $i - d + 1$ additional basis elements $\mathbf{b}_{d+1} \ldots \mathbf{b}_i$ lying in the null space of $\mathbf{B}^{(\mathrm{d})}$ which increase the size of the matrix by augmentation to form $\mathbf{B}^{(\mathrm{i})}$. That is, $\mathbf{B}^{(\mathrm{i})}$ is *inflated* to form $\mathbf{B}^{(\mathrm{d})}$. In this form the reflectance coordinate $\mathbf{x}^{(\mathrm{d})} = (x_1, \ldots x_d)$ in the deflated space maps onto $\mathbf{x}^{(\mathrm{i})} = (x_1, \ldots x_d, 0, \ldots, 0)$ in the inflated space; the transformation matrix $\mathbf{T}$ resembles the identity for its first $d$ diagonal elements. This general form is valid because of the basis orthogonality imposed on all (additional) elements in $\mathbf{B}^{(\mathrm{i})}$.

Preservation of orthogonality has two worthwhile consequences. First, the added elements all lie in the null-space of $\mathbf{B}^{(\mathrm{i})}$ (or its sequence of partial inflations). By this, subsequent deflation restores the basis $\mathbf{B}^{(\mathrm{d})}$ by removing these new basis axes without effecting previously occurring elements, as was illustrated by the mapping of reflectance coordinates above. Thus, deflation serves as inflation's inverse. Since a "blind" restoring deflation omits those axes having least principal component, inflation here must take care to introduce each axes in order of decreasing least-squares contribution. Again, orthogonality is a necessary condition. Second, orthogonality implies independence, so that each new element is guaranteed to increase the rank of the larger basis. Thus, the largest possible inflation stops when $i = m$, as full control of the reflectance function discretized at $m$ points is achieved. That is, the reflectance coordinate $\mathbf{x} = (x_1, \ldots x_n)$ fully determine the reflectance function, as $\mathbf{T}$ is then non-singular and hence of full rank. Since it is orthogonal, $\mathbf{T}$ ultimately describes a rigid rotation and mirroring defining a potentially new coordinate system for the $m$-dimensional reflectance space. When $\mathbf{T}$ is the $n \times n$

identity a useful equivalence between reflectance coordinates and discrete spectral reflectance functions is constructed, as seen in [4.6].

### 7.1.2 Initial space unknown (Case 2)

The second case is both more interesting and more common. It considers the inflation of reflectance coordinates lying within a low dimensional reflectance space of partially-known basis elements. Such reflectance (coordinate) data lying on a partially defined space of representation include, in decreasing level of basis specification, colour scanner data, remote sensing data under conditions of variable illumination and general RGB image data created under unknown photometry. In the first case, the detector profiles utilize the tine model (Figure 2.2) and are known. In the second, data is sensed remotely by detectors having well-defined properties under conditions of varying source of illumination. In the last case the detector responsivities of given RGB image are not known, though general assumptions can be made. In each case the sample may be considered to originate under the action of a hypothetical detector set of partially known responsivity; the spectral power distribution of any source may also be present in implicit form.

In this setting, practical inflation provides a candidate set of known reflectance elements in the larger space to be associated with the reflectance coordinates of unknown origin. That is, a orthogonal frame of reference is created for representation of reflectance data in the inflated space. (Finding either $\mathbf{T}$ or $\mathbf{B}^{(d)}$ establishes the other under known $\mathbf{B}^{(i)}$.) Final choice on a suitable orthogonal transformation $\mathbf{T}$ may be based on any available *a priori* knowledge of what the axes of $\mathbf{B}^{(d)}$ represent.

## 7.2 Analytic Inflation

Inflation now considers the mapping of a largely undefined reflectance space onto a known reflectance space of higher dimension. The analytic treatment employs algebraic and related geometric models concurrently.

The larger, known space is spanned by basis elements which are (fictitious) reflectance functions $\beta_{d+1}(\lambda)\ldots\beta_i(\lambda)$. The set of all physical reflectances $0 \leq \beta(\lambda) \leq 1$ lying within the space defines the locus of valid reflectance coordinates:

$$\mathcal{R}^{(\mathrm{n})} \equiv \{\forall\lambda, \forall(x_1 ... x_i) |\ 0 \leq \Big[\sum_{i=1}^{m} x_i\beta_i(\lambda)\Big] \leq 1\}. \qquad [7.2]$$

This is the reflectance solid $\mathcal{R}^{(\mathrm{i})}$: a convex polytope lying within a Euclidean $i$-space containing all points $(x_1 \ldots x_i)$ solving the conditions of physical reflectance. The algorithmic construction of this solid may be achieved by introducing cutting planes for each wavelength $\lambda$ as discussed in Chapter 4. Symmetry is achieved if the vector $\rho(\lambda) = 1$ lies in the solid's space. When considering nested family of solids, this is possible *iff* $\beta_1(\lambda) = 1$. Thus, a smaller solid $\mathcal{R}^{(\mathrm{d})}$ is fully contained within the parent and lying within a subspace of $n$. For this reason, analytic and empirical methods of inflation may be phrased as problems in geometry.

### 7.2.1 Inflation of 1-Dimensional Reflectance Coordinates

This is the most basic case which is later generalized. The minimal solution considers the inflation of a single reflectance coordinate having unknown photometry on a reflectance space of two dimensions. To simplify notation, the singleton coordinate vector $\mathbf{x}^{(\mathrm{d})} = (x_1^{(\mathrm{d})})$ is replaced by $d_1$, likewise the reflectance vector $\mathbf{x}^{(\mathrm{i})}$ on the larger space will have components $i_1$ and $i_2$. Since any associated reflectance

function must by physical, the axis of inflation must lie within the zonogon $\mathcal{R}^{(i)}$. The search thus considers the set of their intersection: those segments lying within this larger reflectance solid.

Intuitively, the segment coincident with the vector $\beta_1(\lambda)$ is desirable. This grey segment is present in any symmetric reflectance solid and has coordinates $(c, 0, \ldots, 0)$. Geometrically, the segment lies between two cusps (the apses of the solid), giving the span a length which is a local maximum. In all the reflectance solids encountered, it is also a global maximum: printer gamuts invariably exhibit greater dynamic range across the achromatic axis than across any other.

Locating the maximum span is a useful heuristic in situating an axis immune to metamerism in source of illumination and is treated in the next section. Thus, printing this one-dimensional reflectance coordinate $d_1$ on a colour printer will produce media whose associated reflectance functions exhibit (as much as is possible) constant value. This is a "neutral" density plot in the broadest sense. In addition, this 1-D to 2-D inflation provides for a large dynamic range, a consequence of axis length.

Not surprisingly, the preservation of grey balance is both the goal for many monochromatic[1] photographers, and an essential attribute of reprographic *colour correction.* In the latter context, it is termed *END* for *equivalent neutral density* and is a subject of much study (Evans, Hanson and Brewer, 1953). The merit of good grey balance is also evident in production hardware. For example, modern production scanners provide internal colour correction transformations to readjust both for detector responsivities as well as ink impurity. Certain models (Crosfield, 1988) handle near-grey colour correction transformations as a special case to pre-

[1]This title is correctly preferred by many professionals working in "black and white" media.

Figure 7.1: Reflectance Solids in 2-D

serve achromaticity but must interpolate these results against the general model, weighted by proximity to the grey axis, thus removing potential discontinuities or *colour contours* in the device transfer function. The method of local deflation proposed in §6.5 is similar but provides preservation of the achromatic axis by virtue of the underlying "achromatic" reflectance element $\bar{\beta}_1(\lambda)$ common to all local bases.

For completeness sake, there exist zonotopes having two opposing cusps for which the maximum span is not the grey axis. Such reflectance solids would describe printers which can reproduce a wide range of highly saturated colour yet show reduced dynamic range in print density. Two-dimensional zonotopes (zonogons) of both kinds appear in Figure 7.1.

In the all solids studied, only form (a) was encountered. Procedurally, using this axis without further consideration assumes there no overriding thematic meaning, such as using "red" pseudo-colour to indicate an infra-red thermal reflectance channel. The choice of this grey axis is justified below.

### 7.2.2 Metamerism Considerations

Under an unknown source $\Phi_\lambda$ there exist lines of metameric colour when the data on $\mathbf{B}^{(\mathrm{d})}$ is inflated onto $\mathbf{B}^{(\mathrm{i})}$. Points along this line describe distinct spectral reflectance functions having identical coordinates in the deflated space. An inflation axis which parallels the metameric lines of common orientation is clearly undesirable, for a change in value ($x_1^{(\mathrm{d})}$) would then not produce any perceptual change under a selected source of illumination.

When the axis of inflation describes a non-negative basis vector, increasing this parameter increases the overall value of the reflectance function. However, orthogonal reflectance functions showing both positive and negative excursions may have little or no contributing effect to the profile of a spectral reflectance function under some source of illumination. The specific case of "black" reflectances lying within the null space associated with a specific source of illumination has been explored (Wyszecki and Styles, 1982d). In the case at hand, the condition is explored generally under an unknown and indefinite source $\Phi_\lambda$ (Brainard, Wandell and Cowan, 1989).

The problem may be expressed algebraically in reference to a parameterized source. The best axis then has an orientation making it the least sensitive to metamerism across a universe of all possible sources under the assumption that all sources are equally likely. The unknown inflation may be considered a deflation (detection) from the larger space of representation which establishes the lone reflectance coordinate $d_1$ lying within the deflated reflectance space spanned by $\beta_1^{(\mathrm{d})}$. Mapping reflectance coordinates onto a spectral reflectance and then deflating gives

$$d_1 = \int [(i_1 \bar{\beta}_1^{(\mathrm{i})} + i_2 \tilde{\beta}_2^{(\mathrm{i})}) \Phi_\lambda] \; \tilde{\beta}_1^{(\mathrm{d})} d\lambda. \qquad [7.3]$$

Holding the detector response $d_1$, fixed, the locus of all points $\mathbf{x}^{(\mathrm{i})} = (i_1, i_2)$ sat-

Figure 7.2: Line (segments) of Metamerism Under Change of Illumination

isfying the equation define metameric reflectances in the larger space under conditions of source and detector. (For purposes of simplification, any perceptual colour matching function such $\bar{x}(\lambda)$ may be associated with the detector responsivity present within the integrands.) The equation may be rewritten as:

$$d_1 = i_1 \int (\bar{\beta}_1^{(\mathrm{i})} \tilde{\beta}_1^{(\mathrm{d})} \Phi_\lambda d\lambda + i_2 \int (\bar{\beta}_2^{(\mathrm{i})} \tilde{\beta}_1^{(\mathrm{d})} \Phi_\lambda d\lambda. \qquad [7.4]$$

or

$$i_2 = \Big[\frac{-\beta_1^{(\mathrm{i})}(\lambda)\beta_1^{(\mathrm{d})}(\lambda)}{\beta_2^{(\mathrm{i})}(\lambda)\beta_1^{(\mathrm{A})}(\lambda)}\Big] i_1 + \Big[\frac{1}{\beta_2^{(\mathrm{i})}(\lambda)\beta_1^{(\mathrm{d})}(\lambda)}\Big] d_1. \qquad [7.5]$$

This is a line $y = mx + b$ in slope-intercept form. Here the Cartesian axes are the orthogonal bases $\beta_1^{(\mathrm{i})}$ and $\beta_2^{(\mathrm{i})}$ of the inflated space. The slope and intercept are functions of the unknown but invariant detector response $\beta_1^{(\mathrm{d})}$ and parameter $\Phi_\lambda$. Holding the latter constant describes lines of metamerism under a fixed source of illumination. These are segments within the defining region of the zonogon $\mathcal{Z}^2$. The intercept of each varies with reflectance coordinate $d_1$, though the slopes are constant for fixed source. This is illustrated in Figure 7.2.

Thus, a range of increasing reflectance coordinates $d_1$ indexes points along successive segments. The point on each line is indeterminate; metamerism indicates that for the given illuminant they produce indistinguishable results. Returning

Figure 7.3: Inflation Paths Across Segments of Metamerism

to this section's primary concern: a desirable inflation should not parallel any metameric line. The trial inflation paths presented in Figure 7.3 satisfy this for the underlying source, but are undesirable.

In (a), the path is two circular (quadratic) arcs chosen to approximate the set of midpoints of each metameric line segment. Such a path is an impossible candidate for defining an inflationary path, as it is not linear. In (b) a path is chosen which crosses all metameric lines at right angles (not immediately evident given the plane figure's illusion of perspective), thereby maximizing the ascent of the gradient of change. However, this path does not cross all metameric segments associated with the defining source. Neither does it begin at the origin: representation of the coordinate $d_1 =$ in the inflated space would not be the black reflectance.

Preservation of the black reflectance under inflation may be easily achieved. Any inflation of the reflectance coordinate $d_1 = 0$ creates metamers intersecting at a common point $i_1 = 0$, $i_2 = 0$ made clear by substitution into the slope-intercept equation [7.4]. This formally establishes the origin as the logical starting point. An inflationary path coincident with the grey axis guarantees a span having (locally) maximal length, thereby fixing the path's slope. Figure 7.4 depicts the metameric segments for three distinct sources. All are coincident at the origin.

Figure 7.4: Orientation of Segments of Metamerism

Although intuition would suggest that an inflationary axis crossing metameric lines at right angles is "best", it is the number of lines crossed which determine the overall change in detector value. Line crossings, not orientation or position best defines a "metameric gradient". Thus, the choice is the reuse of the inital axis $\beta_1^{(i)}$ as in practice it spans the conjugate vertices most removed in the reflectance solid.

A final consideration is the range of sources of illumination for which this axes remains non-metameric. Assuming that an unknown set of sources produces orientations which are uniform by angle, then that axis about which the greatest number of (non-parallel) lines intersect is the most desirable. This inflation axis is least likely to introduce metamerism. This may be verified by applying the principle of a "line of support" borrowed from the domain of solid geometry and demonstrated in the next section.

### 7.2.3 Lines of Support

A geometric method is now introduced to locate the longest (set of) axes. The best single axis cuts the greatest number of parallel lines in an "sheaf" regardless of that sheaf's orientation. (Orientation is a function of the source of illumination's spectral power distribution). As suggested by figure Figure 7.5, this axis appears to run

Figure 7.5: Points of Contact of a Rotating Polygon

between those vertices having the smallest angle of subtense. Formalizing this claim and locating the most remote vertex may be accomplished through introduction of a geometric device. Consider instead that the "sheaf" remains at a constant orientation while the reflectance solid rolls. That is, it rotates while keeping its lowest point (in $y$ extent) in contact with the $x$ axis and its highest point in contact with a parallel plate. That is, the figure is a cam; the point of highest position then found by a cam follower. The greatest $y$ extent occurs for the set of vertex pairs which remain in contact with both the base and the plate throughout its greatest angular rotation. An example appears in Figure 7.5 in which the accompanying circular diagram defining the pairs of vertices of upper and lower incidence as a function of circumferential angle.

The vertex pair $(A, C)$ remain unchanged during rolling for the longest time. The associated span cuts the fixed metameric shear through the greatest period of angular rotation – the desired property. This derivation follows the "line of support" model familiar to descriptive solid geometry (Grünbaum, 1967) whose methods of general solution are well-known. The desired vertex pair is most often those of the extreme vertices having associated acute angles. Thus, the grey inflation axis also

locates the path of greatest span, hence greatest insensitivity to metamerism under arbitrary source. Geometrically, the path spanning the two vertices holding contact under the longest change in angular subtense are desired.

Generalization of span to higher dimensions is difficult, as rotation about multiple axes is not fully ordered. With zonotope reflectance solids, both the base and plate always make contact with opposing pairs of vertices and the maximum distance occurs through a unique zone (all lines of extent pass through the figure's centroid, though the $y$-position of the centroid remains variable). When inflating to high dimensions, broadest (solid) angular subtense may again be approximated by longest span. The latter may be determined exactly using maximum spanning tree and related graph-theoretic algorithms, which locate most extreme vertices in near-linear time (Aho, Hopcroft and Ullman, 1983). Unpublished methods (Heckbert, 1985) suggest that many zonotope-related algorithms may run in time $O(n \log n)$. More recently, abstract geometers (Bouma and Vaněček Jr., 1993) have refined object intersection models to simulate physical contact of objects, of particular use in CAM.

### 7.2.4 $n$-Dimensional Inflation with Unspecified Source

Geometric methods and not graph-theoretic ones are well-suited to generalizing the principle of maximum span. For instance, the inflation $\mathcal{R}^2 \stackrel{\mathrm{I}}{\rightarrow} \mathcal{R}^3$ locates a cross-sectioning plane of maximum area within the polyhedron, as this maximizes the "flux" of penetrating metameric segments across all possible orientations. Unfortunately, computational geometers cannot yet provide general statements regarding the metrics of such cross-sections, though rigorous proofs for a number of non-

Figure 7.6: Maximum Cross-section of the Cube

intuitive situations are known[2].

Other approaches, such as generalization of the $1-D$ case by application the "greedy" method to locate successive axes may not work. As a counterexample, Figure 7.6 shows a cube (the prototypical zonotope) in $\mathcal{R}^3$ having a cross-section of maximum area which contains the axis of greatest $1-D$ length (the body diagonal) yet contains no axis normal to it.

Nontheless, cutting planes which satisfy either property may make good candidates for inflation. In fact, to be consistent with formal properties deflation, the planar cross-section would ideally contain the body diagonal. This is formalized below.

### 7.2.5 Property of Nested Containment

Successive dimensional inflation should provide for "nested containment" of reflectance solids. This constraint makes inflation transitive along a set of reflectance

[2]For instance, a solid of (isotropically) invariant perimeter has invariant cross-section, yet need not even remotely resemble the sphere.

solids $\mathcal{R}^1 \ldots \mathcal{R}^n$ of increasing dimension, by analogy to [6.3]:

$$\mathcal{R}^m \overset{\mathrm{I}}{\rightarrow} \mathcal{R}^n \equiv \mathcal{R}^m \overset{\mathrm{I}}{\rightarrow} \mathcal{R}^{m+1} \overset{\mathrm{I}}{\rightarrow} \ldots \overset{\mathrm{I}}{\rightarrow} \mathcal{R}^n. \qquad [7.6]$$

It therefore defines a full ordering, by analogy to [5.14]:

$$\mathcal{R}^{n-1} \subseteq \mathcal{R}^n \subseteq \mathcal{R}^{n+1}. \qquad [7.7]$$

Dimensional containment assures that a deflationary "return path" exists to the higher dimensional antecedent. Since inflation increments the dimension of inflation, all inflations (potentially showing dissimilar span despite similar dimension) necessarily converge in the common $m$-dimensional reflectance space of maximum discrete precision. In this fashion, inflation carries representations upward within the Colour Hierarchy to a point of common definition. Deflation then allows descent to an alternative representation.

The ultimate path need not be unique. For instance, greedy methods proceed deterministically and thus define a full ordering, yet do not necessarily discover intermediate spaces of best representative power while approaching the root of the colour tree. Instead, the limiting the number of potential axes to manageable amounts suggests a solution by interactive geometric tools. This allows the identification of best inflation axes by problem-specific concerns, especially where deterministic methods are left with possible axes of near identical span yet distinct orientation.

## 7.3 Interactive Inflation

### 7.3.1 Tools Overview

The geometric study of inflation requires visualization aids. The interactive modeler, whose primary window depicts convex reflectance solids in perspective and true colour, is extended. A secondary window provides a simultaneous view of planar cross-sections of the solid. These are rendered in orthographic projection and annotated. Tertiary windows plot spectral reflectance functions or power distributions, either as graphs or on CIE-defined colour charts. A screen snap-shot of the system with all windows active appears in Figure 7.7. (The solid's symmetry is not pronounced owing to the forced perspective.)

Because the cross-sections are two-dimensional, the modeler's greatest value lies in locating a secondary axis of inflation by interactive means. (The primary axis is the initial constant grey reflectance forming the "body diagonal" for analytic and practical reasons previously set forth.)

### 7.3.2 Reflectance Cross-sections

The interactive system works as follows. The parent reflectance solid is oriented using the rotation/reference frame tools within the parent window. Conceptually, the solid is best considered to lie "half in" the faceplate of the display, thus containing any reflectance solid's centroid of 50% grey. Invocation of the cross-section tool introduces a cutting plane containing the display $x_1$ axis which corresponds to the initial $\bar{\beta}_1(\lambda)$ grey basis vector. Determination of the second axis describes a rotation about the primary axis $x_1$ and fully determines the transformation $\mathbf{T}$ appearing in [7.1]. Convexity guarantees that any axis of (local) greatest extent

Figure 7.7: The Interactive Solid Model

Figure 7.8: Reflectance Cross-Sections of Cone and MSSM Solids

spans opposing vertices and passes through the centroid. The solid is then rotated through a discrete angle until a vertex set coincides with the plane of the display. Repeated invocation of the command locates the next (previous) set of candidate vertices. This tool thus provides a partial solution in finding the best axes by reducing the search to a discrete set of choices. Cross-sections of both the reflectance cone and of the MSSM-derived Kodak printer reflectance solid appear in Figure 7.8. The solid's true colour may be depicted on the Iris display under action of source E. By way of reference, the reflectance function associated with the peak apex in (a) has chromaticity coordinates $x' = .523$, $y' = .448$.

The $2 - D$ space of inflation is thus determined by the orientation of the parent solid within the plane. The convex hull of the two axes (intervals) is a parallelogram, depicted under orthographic projection in a companion window. Its surface is rendered in exact colour (each pixel in perfect spatial agreement with the tristimulus vector it describes) under any given source by virtue of simple linear interpolation. This is provided directly on the SGI/Iris: merely the RGB values of the white and black apses and the antipodal extreme points need be passed. These are in linear

relation with the reflectance coordinates at any point. Moreover, linear interpolation of these chromaticities between vertices correctly describes the reflectance coordinates of the section at that point. This demonstrates the utility of linear colour spaces in relation to (linear) spaces of world and display spatial coordinates. Their interrelation was taken up in §6.3.3 (second subsection on hardware rendering and linearity).

### 7.3.3 Opponent Reflectances

Note that each parallelogram describes a set of complementary reflectances. The centre of each figure is the point of 50% grey reflectance. Any two endpoints $a$ and $b$ of a line segment equidistance from this centre have reflectance functions $\rho_a(\lambda)$ and $\rho_b(\lambda)$. As above, agreement between spatial and colorimetric linearity states that the average reflectance value of each is the reflectance value of the averaged (middle) point. That is,

$$\frac{1}{2}[\rho_a(\lambda) + \rho_b(\lambda)] = \frac{1}{2} \qquad [7.8]$$

so that

$$\rho_a(\lambda) = 1 - \rho_b(\lambda). \qquad [7.9]$$

These are the *opponent reflectances* defined in §5.2.2. From [7.9] it is immediately clear that $\rho_b(\lambda) = 1 - \rho_a(\lambda)$ also: opponents occur in pairs, save for the midpoint $\rho_{mid}(\lambda) = \frac{1}{2}$ which is self-complementary.

The smooth hue gradient and range of opponent colours appearing in the hardcopy should not be taken for granted. In general, the first-principles creation of systems of complementary colour seldom gives rise to perceptually useful complements. As an illustration, the hue of a source of illumination $c$ may be matched

to hue of some monochromatic source with emissivity at $\lambda_c$, yet, the wavelength of a complementary $c'$ is not easily determined, nor necessarily spectral (green and purple are complements). Moreover, a linear plot of colour of a progression of monochromatic sources does not provide remotely uniform widths for certain hues, the "yellow" range being particularly narrow.

Complementary colours are useful in the additive domain in creating visually distinct colour pairs, as when selecting text regions. A method pair production based upon reversible interval swapping has been advanced (Paeth, 1992) extending a specific framebuffer technique (Higgins and Booth, 1986). Other solutions based upon geometry are known (Turkowski, 1986), (Lai, 1991). In the reflective domain, the use of a simple mathematical model supporting a related geometric inversion is of particular value (see also §8.3.2).

### 7.3.4 Cross-sections and Parallelograms

The true-colour parallelogram is similar but not necessarily identical to a cross-section of the 3D reflectance solid by the plane of the display. In general, the cross-section of a convex solid need not be identical to its *silhouette* (image under orthographic projection), also called a shadow (McKenna and Seidel, 1985). In all cases, the latter is the projection of the convex hull and thus contains the former, here convex containment is preserved under orthographic projection. This is particularly evident when a cross-section having different orientation is taken.

As seen in Figure 7.9, the cross-section of the reflectance solid (cube) in (a) reveals a hexagon[3]; in (b) the silhouette includes the projection of the six vertices (all, save for the black and white apses), which are not present in (a). The figures

[3]The cross-section of a zonotope remains a zonotope (Coxeter, 1973).

Figure 7.9: Cross-Section versus Silhouette

lie in closer agreement should the cycle of edges connecting these vertices lie closer to planarity. This is taken up again in §7.4.1 (see also Figure 7.11).

### 7.3.5 Sheaf Models

The vertices lying about the belt occur in opposing position (and colour), along an axis passing through the centroid. The black and white vertices share this property, defining a second, independent axis through the common grey centroid. Thus, a finite set of parallelogons may be constructed which are indexed by successive vertices through one half-rotation of the reflectance solid. The set of reflectance solids thus produced are called a *sheaf.* The sheaf of parallelograms thus enumerates all vertices and hence is a discrete model of the convex hull of the reflectance solid. The sheaf may be produced using the modeler by reactivating the "cross-section" menu item, thus rotating the solid until the next belt vertex lies in the plane of the display. (The display window is annotated with both turning angle and surface

Figure 7.10: Sheaf Cross-sections (Kodak SVC6500 Reflectance Solid)

area.) Two additional parallelogons (compare Figure 7.8(b)) having distinct angle and culled from the set of nineteen parallelogons forming the SVC6500 reflectance sheaf are reproduced in Figure 7.10.

The grey axis of rotation common to each parallelogon suggests creating a three-dimensional physical model having leaves interpenetrating all others along this axis. This is analogous to the physical depiction of trichromatic colour spaces using (3D) physical models in which the leaves are hinged panels of varying hue hung from an upright "grey axis" post, or like the Munsell colour model (Munsell, 1963). In the present case, the neutral $x_1$ axis runs horizontally and is a reflective grey in a colour space in the broadest sense: it is grey without reference to models of illumination or perception.

## 7.4 Reflectance Belts

A simplified representation of any reflectance solid records merely[4] the belt. The simplified solid is then reconstructed as the convex hull of the belt and the black and white apses. The approximate solid thus defined is fully contained by the original reflectance solid. Its definition is conservative: the solid contains no regions of non-physical colour. The enclosed volume matches the parent solid when all vertices lie within a belt; the difference in volume goes to zero as any non-belt vertices lie in closer proximity to the belt. As the number of vertices increase with sampling resolution, the belt becomes a continuous curve which describes the extreme girth of a continuous reflectance solid. Continuity of slope both along and across the belt are described in §4.2.9.

The belt describes a set of highly saturated colours which plays an analogous role to the spectrum locus (convex hull) of the CIE chromaticity diagram. Both are convex sets in which the colorimetric coordinates of intermediate "blended" colour (by respective superposition or composition) lies along the segment joining the coordinates which represent the original sources. In both cases, this locus defines the interior. Convex closure of the additive model introduces a "purple line" joining the extreme ranges of spectral sources (and is thus non-spectral). An analogous artifact does not exist with the reflectance solid.

This suggests that the belt is not merely a "device" useful in forming a compact model, but a geometric feature which encodes many of the properties of reflectance solids. The nature of the MSSM belt is studied below; it is generalized to higher dimension in §7.5.

[4] In practice, only half the belt's perimeter need be recorded.

Figure 7.11: Belt Projection versus Cross-section

### 7.4.1 Planar Belt Approximations

Belt planarity is a valuable property, for it allows the complete representation of a $n$-dimensional reflectance solid using an $(n-1)$-dimensional (convex polytope) model. Typical belts deviate from (hyper)planarity but can be approximated. Unfortunately, parallel projection of the belt vertices onto the plane, normal to $x_1$ through the centroid, potentially introduces vertices lying outside the reflectance solid: the approximating reflectance solid may then describe points of non-reflective colour. This is made clear by considering the projection of belt vertices of the basic reflectance cube appearing in Figure 7.11(a) onto the cutting plane which forms the cross-section in (b), demonstrating containment in (c).

The proper solution is to consider merely this reflectance cross-sectioning as in (b), thus assuring containment by the parent solid. As can be seen in the figures, central sections of zonotopes remain zonotopes (Coxeter, 1973); here the cross-section of the cube yields the regular hexagon[5]. Thus, the planar belt formed in this fashion has the added advantage of also being a "reflectance solid" yet which is curiously devoid of luminance information (the axis of projection).

As a further example, the four-dimension reflectance solid based on the Kodak data may be represented as a three-dimensional polyhedron of projection whose sur-

[5]By analogy, the central section of the 4D hypercube yields the 3D regular octahedron.

face defines those reflectance functions lying near the limits of the physical condition [2.3] at four wavelengths and meeting these limits exactly wherever the surface of the approximating solid also intersects this plane of projection. Solids such as these approximate the *four*-dimensional belt using a *three*-dimensional model. Higher dimensional models are described again in §7.5.

When a loss of symmetry is permissible, the sectioning plane through the centroid may be reoriented to provide a better least-squares fit of the belt vertices. See in particular Figure 4.4.

### 7.4.2 The Unbiased Belt

The belt showing no bias must be both planar and of constant $y$ height. Thus, it is circular. The reflectance solid defined strictly by this belt and two apses is presented below.

The circular belt of unit extent coaxial about the unit interval $0 \leq x_1 \leq 1$ is the locus of points

$$[\frac{1}{2},\ \cos 2\pi\theta,\ \sin 2\pi\theta] \qquad [7.10]$$

in which $\theta$ gives the anti-clockwise (viewed from the origin) angular bearing in circumferential units about $x_1$ of any point on the belt's perimeter. Since the belt lies on the solid's service it satisfies the limits of the reflectance inequality:

$$x_1 + x_2\beta_2(\lambda) + x_3\beta_3(\lambda) = 1, \qquad [7.11]$$

$$x_1 + x_2\beta_2(\lambda) + x_3\beta_3(\lambda) = 0 \qquad [7.12]$$

in which the unit factor term $\beta_1(\lambda) = 1$ has been removed. Substitution of [7.10] into [7.12] gives

$$[\cos 2\pi\theta]\ \beta_2(\lambda) + [\sin 2\pi\theta]\ \beta_3(\lambda) = \pm\frac{1}{2} \qquad [7.13]$$

which must be satisfied with suitable choice of orthogonal functions $\beta_2(\lambda)$ and $\beta_3(\lambda)$ valid for any $\theta$.

Recalling the identity $\cos^2\alpha + \sin^2\alpha = 1$ for all $\alpha$, then

$$\beta_2(\lambda) = \frac{1}{2}\cos 2\pi\lambda', \qquad [7.14]$$

$$\beta_3(\lambda) = \pm\frac{1}{2}\sin 2\pi\lambda' \qquad [7.15]$$

provide suitable choices. Thus, given an invariant belt and the reflectance solid symmetry condition $\bar{\beta}_1(\lambda) = 1$, two orthogonal sinusoid basis elements are derived. In [5.3.3] these basis elements were presented *a priori* and give rise to the bi-conical model appearing in Figure 5.5. Here the converse is proved: the convex hull formed by this planar circuit and two apses uniquely determines the sinusoid basis.

This solid shows no best inflationary axis: all vertices along the belt are perpendicular to the grey axis at the centroid; their distance to this point (hence, half their total span) is constant.

### 7.4.3 Biased Belts: Non-Planarity and Excursion

Since belt need not be planar (unlike the locus of spectral colour); this condition is not required for the belt to establish the convex hull exactly. For empirically derived data, excursions from planarity may be plotted by recording the location of the belt vertices on an $yz$ plane through a half-rotation of the $x$ axis. (The Kodak MSSM data provided a total of thirty-eight belt vertices.) In Figure 7.12, nineteen consecutive vertices are plotted by half-rotation of the reflectance solid about the $x$ axis. The colour of each vertex is computed under source E; the wandering of the projected points about the centre line reveals both deviation from planarity

Figure 7.12: Belt Eccentricity "Baton Sweep"

(motion along the $x$ axis) and total excursion away from the monochromatic axis ($y$ height).

### 7.4.4 Choices for Inflation

Note that the belt traverse reveals both local and global extrema about the circuit. A maximum showing largest departure suggests its choice as the axis which best completes the inflation $\mathcal{R}^2 \stackrel{\mathrm{I}}{\rightarrow} \mathcal{R}^3$. For the data presented here, this further reduces the number of viable choices of inflation from nineteen to three. Note that the global maximum is also that vertex lying closest to the central line, thus lying in the plane normal to the axis $x_1$ through the reflectance solid's grey centroid.

### 7.4.5 Belt Traversal

The belt vertices may be traversed cyclically. Since the belt dihedral is defined by the intersection of the black-apex and white-apex cone, the belt edges define limiting reflectances satisfying both $\rho(\lambda') = 0$ and $\rho(\lambda'') = 1$. At each consecutive vertex (feature of dimension zero) contact is made with a third limiting zone. Most often, this vertex is a sampling artifact (§4.2.9) lying at the discrete wavelength

adjacent to either $\lambda'$ or $\lambda''$. In other cases, a third, distinct minimum or maximum well-removed from the other two is encountered. Consequently, the belt undergoes a slope discontinuity and bends, forming a *joint.*

Such a vertex introduces a prominent face to the surface of the solid, which cannot be sculpted by finer sampling. The partitioning zone containing this face reduces the volume of the solid as compared to a gradual transition. The portion lost is the set of points whose reflectance functions exhibit a peak wavelength between the extremes. Put another way, the occurrence of the the second matching extremum limits the range of values which the reflectance function could otherwise take on along the domain defined by the two like-signed extrema.

Most often, two creases exist in the reflectance solid, somewhat analogous to the creases present where the curving spectrum locus turns abruptly to form the linear purple line. In this case, the conditions of monochromatic source become the potential discontinuity when the reflectance function on the limits of extreme wavelength $\rho(\lambda_{max})$ and $\rho(\lambda_{min})$ are considered. If the values agree (as with the bi-conical model), the $C^0$ continuous surface of the solid will additionally be $C^1$ continuous. Other possible creases along the belt suggest further study of empirical data.

### 7.4.6 Limit Codes

A limit code is a reduced description of the minima and maxima occurring at any vertex on the surface of the reflectance solid. Reflectance coordinates of the belt vertices are discarded together with the wavelengths where $\rho(\lambda)$ achieves its extrema. Regained are merely the order and "sign" of the extrema, with "1" indicating upper zone and "0" the lower. Since any vertex is common to the intersection of $m$ zones,

an $m$ digit code word records these belt limits for a solid in $m$-space. The digit string is termed a *limit code.*

Limit codes are useful in locating joints (dihedral breaks) along the belt. By arguments of continuity and convexity, these vertices necessarily maximize the distance from the solid's centroid. Thus, they are the same vertices as the local and global maxima appearing in the "baton sweep" plot in Figure 7.12. While a relative run length might be useful indicating dominant limit code, a means of flagging joints along the spectral belt is more useful. As mentioned in §7.4.4, this information may be used in either automatic or interactive fashion, thus finally determining an appropriate axis, and hence related transformation, which fully defines inflation. The codes of two prototypical models, the bi-cone and the cube, may now be studied.

For motion along the sinusoid's belt, the extrema remain at a distance fixed by the one quarter of the interval of wavelength (ninety degree quadrature). Continuous adjustment of phase reveals no discontinuities at the ends of the visible spectrum. However, the passing of the minimum off the far red (maximum wavelength) end of the spectrum and its simultaneous reintroduction at the far blue (minimum wavelength) end complements the limit code at that point. A canonical set may be formed by considering codes beginning with a "0" for some turning angle and continuing through 180°. Note that for any solid the limit code of a conjugate vertex is easily found by complementing the code's digits, made clear by the mapping of reflectance functions about the centroid shown in the proof outline following [5.22]. (See also Figure 8.11.)

The limit codes of the (3-space) cube are more interesting. The three axes are the three diagonal reflectance functions of non-overlapping unit area at low, medium and high wavelength. By this, the coordinates of the unit cube with first vertex coordinates $(\frac{1}{2} \pm \frac{1}{2}, \frac{1}{2} \pm \frac{1}{2}, \frac{1}{2} \pm \frac{1}{2})$ are identical to those vertices limit codes.

Figure 7.13: Cube Belt Traverse (non-planar hexagon)

The belt traverse with codes (vertex coordinates) appears in Figure 7.13.

The traverse is non-planar hexagon which visits all vertices save the achromatic apses; it is the Hamiltonian Circuit (Aho, Hopcroft and Ullman, 1983) on a cube having two opposing vertices removed. Since (rectilinear) movement on any axis changes merely the coordinate for that axis, the cycle of vertices produce values which form a *Gray code*[6] (Ludman and Sampson, 1981) of partial length, as code words 000 and 111 are omitted. All cube traverses share this Gray code property. Ignoring rotations and mirroring, the limit code for the cube as appearing in Figure 7.13 is

$$\{ \quad 001 \quad 011 \quad 010 \quad 110 \quad 100 \quad 101 \quad \}. \qquad [7.16]$$

This list suggests a left (right) shift of the minima (maxima) as the circuit is traversed. In fact, imaging a feature 1.5 pixels wide across a three-pixel scan line will reproduce a picket fence discretization matching the list above. This prototypical model may be used as a starting point in analyzing empirically derived reflectance solid data.

[6]This is a $2^n$ length sequence of $n$-bit binary codes in which any two adjacent codes have a *Hamming distance* of one, that is, differ in exactly one bit position.

| Code | Length | Angle | Area (@ angle) |
|---|---|---|---|
| 011 | 1 | 2.1° | 28.6 |
| 110 | 7 | 2.5° | 29.0 @ 5.9° |
| 011 | 1 | 16.3° | 27.9 |
| 110 | 1 | 21.0° | 27.0 |
| 001 | 3 | 22.4° | 26.8 @ 22.4° |
| 101 | 1 | 66.0° | 25.8 |
| 110 | 1 | 85.1° | 23.1 |
| 100 | 1 | 92.5° | 23.3 |
| 110 | 2 | 96.0° | 22.3 @ 104.7° |
| 100 | 5 | 119.7° | 26.3 |
| 110 | 2 | 136.5° | 26.3 @ 136.5° |
| 100 | 2 | 156.3° | 26.3 @ 168.0° |
| 110 | 1 | 174.1° | 27.2 |
| 100 | 1 | 177.9° | 28.6 |

Table 7.1: Empirical Limit Codes and Run Lengths

### 7.4.7 Limit Codes in Practice

Analysis of the belt vertices for the Kodak SVC6500 data set yields a set of limit codes. These are presented in Table 7.1. Runs of adjacent codes have been compacted, leaving the starting (vertex) angle of the run. In addition, the area of the maximum parallelogon cross-section occuring within the run is presented. The area value is in linear proportion to distance between the belt vertex and the central $x_1$ axis.

Only half the table is reproduced owing to its complementary nature; the first and last entries of the table might be considered opposing entries of the same run.

The reflectances corresponding to the runs represent, in sequence, a high-pass (red) function in which the lower wavelength end raises and sweeps its maximum toward the longer wavelength. (That is, the hues of the corresponding belt vertices are moving from blue toward red when taken in order of increasing angle.) The plot of maximum area occurs early on 5.9° (its cross-section appears as Figure 7.8(b)); further turning reveals local and global minima, also tabulated. Codes in general show a bit-shifting pattern analogous to the $2n$ code words produced on a $n$-bit Johnson ring counter. Of particular interest table entry six: the limit code 101 (which together with 101 is one of two forbidden counter codes). Its cross-section appears as Figure 7.10(a); its related reflectance shows a pronounced notch as the extrema are not adjacent. The complementary entry 010 nearly occurs at an angle of 110° as seen in Figure 8.11(c), though the right-most curve edge does not touch down.

Limit codes are most useful when they establish persistent vertex features. The overall length of Table 7.1 is increased unnecessarily because noise present in the nearly flat portions of the red and magenta reflectance spectra (Figure 6.7) can cause multiple touch-downs and lift-offs, breaking up otherwise long runs. Thus, some smoothing of data is useful to insure that the limit codes represent merely *features* and not sampling artifacts along the belt.

## 7.5 Solids in Higher Dimensions

Since visualization of higher-dimensional solids is nearly an impossible[7] task, methods of representation in lower-dimensional spaces are required (Noll, 1967) (Gunn,

[7] Alica Boole Stott was remarkably skilled in 4-D visualization of polytopes and assisted prominent geometers into the 1930's by describing their construction (Coxeter, 1992).

1993). Although projection – the very process by which the solids are themselves created – seems one route, cross-sectioning was employed instead. This approach avoids the algorithmic identification and categorization of extended geometric features (facets), already difficult in 3-space. In fact, the vertices of the 4-dimensional reflectance solid were never enumerated, though this is easily done using methods surveyed elsewhere (Pulleyblank, 1989). Instead, a progressive set of cross-sections are used to create three-dimensional "cutaway" views using methods outlined below.

Note that an $\Re^3$ central cross-section of the $\Re^4$ solid will closely approximate the convex hull of the $\Re^4$ solid, provided the latter's belt is close to (hyper) planarity, after §7.4.1. At the present, the departure from planarity (and general nature) of the belt in dimensions higher than three is unknown as no vertex enumeration has been conducted. This is clearly an area of further study.

The sheaf models of §7.3.5 are easily generalized to the fourth dimension. The method's extension (not pursued) considers a sequence of similar parallelohedra all having both the grey and a secondary axis in common. These rotate through an angle about the orthogonal axis representing the remaining degree of freedom. Ideally, a hue angle exists by analogy to the methods of sheaf choice, previously described. However, one is faced with a hyperplane rotation about two axes; there are no defining criteria for the secondary turning axis. Thus, a succession of solids of changing colour yet otherwise having similar bi-conical shape can reveal the nature of neither its four-dimensional reflectance solid nor its belt.

### 7.5.1 Cross-Sections in 3-D

Instead, a sequence of cutting planes is used to provide a set of representative reflectance cross-sections in $\Re^3$. Methods of *constructive solid geometry (CSG)* are helpful here. Generally, the *sweep* of a solid of lower dimension (Pegna, 1987), (Weld, 1987) may be used to define a solid of higher dimension. Most often, these are used to define (3D) volumes by the motion of a (2D) defining surface of potentially variable shape along a trajectory. That is, they formalize the notion of a set of $2\frac{1}{2}$D "cutaway" drawings, as to a draftsman. Recently, *extrusion* methods have generalized sweep models (Ferrucci, 1993). For present purposes, the construction of 3D physical models is not important: the path of generalization here lies toward simpler linear motions but at increased dimension, explored elswhere (Ferrucci and Paoluzzi, 1991).

A cutting axis must be chosen. As mentioned before, a central plane normal to $x_1$ reveals a solid closely approximating the belt: its surface describes reflectances having four extrema. Of these, non-adjacent double maxima and and minima describe the most highly saturated colours possible in the reflectance space. The construction of this solid is the subject of a follow-on paper. Unfortunately, its neighboring cross-sections are not illustrative of the reflectance solid. By analogy, a cross-sections of the cube across the body diagonal create a sequence of triangles, distorted hexagons and finally a regular hexagon at the midpoint, then reverse[8]. Though highly symmetric and of use in revealing the central hue cycle (compare Figures 5.2 and 5.4), the cross-sections are not remotely cubelike. By comparison, the cross-sectioning of a cube along any coordinate axis forms a sequence of squares. This suggests a set of $3D$ sheaf volumes sectioned along a coordinate axis and

[8]Coxeter depicts such 4D cutaways in a comprehensive appendix (Coxeter, 1973).

Figure 7.14: A $3\frac{1}{2}$D cutaway of the 4D Reflectance Solid

presented in sequence.

A set of cross-sections normal to the $x_4$ plane was thus created. The progression of solids therefore resemble the original reflectance solid in $\Re^3$ appearing in Figure 6.9, but with surface colour and faces changing slowly as the solid progresses along the axis of higher dimension[9]. Note that solids in proximity to the black apex necessarily are dark, as their surface defines reflectance functions having multiple minima. The solids appear in right to left, top to bottom order in Figure 7.14.

The cross-section of the first solid (upper right) is displaced inward from the black apex, else merely a point would be formed. The sections continue until

[9]The modeler allows this sequence to be viewed in kinetoscope fashion.

Figure 7.15: Chromaticities of the 4D Reflectance Solid

the midpoint is reached. The remaining half of the sequence depicts solids of reversed handedness[10] with surfaces define complementary reflectances, owing to central inversion[11]. That is, there reflectance functions are conjugates of the type appearing later in Figure 8.11.

### 7.5.2 Chromaticities of the 4-D MSSM Solid

The chromaticity of all vertices comprising the solid in $\Re^4$ may be plotted by reusing the modeler's tertiary windows. Two chromaticity digrams were created and appear in Figure 7.15. The chromaticity of the additional element $x_4$ is a complementary green (plotted as the numeral 4 in Figure 6.21), increasing the size of the gamut toward the diagram's green apex.

The low-density plot appearing in (a) was created from four cross-sections along the $x_4$ axis having relative offsets of $\{.2,\ .4,\ .6,\ .8\}$. In (b) nine plots were created

[10]This is mirroring in four dimensions, as through a Klein bottle.

[11]Lacking a canonical form for $4D$ vertex enumeration, the custom modeler occasionally produces a solid of reversed handedness along the sequence.

from cross-sections of the same solid in which the interval distance has been halved to .1, thereby filling in[12] more detail. An extended study of the elliptical strands of (b) suggests a cuplike, continuous surface open along the top. This suggests that the points would be better plotted in 3-dimensional tristimulus space and viewed using interactive controls of orientation. This is easily accomodated by the Iris hardware and presents another avenue of further study.

## 7.6 Summary

Inflation provides a means of representing reflectance solids within a space of increased dimension. A practical application is the accurate rendering of monochromatic image data of unknown or vaguely defined photometry on a digital colour printer whose controls provide additional degrees of freedom.

Because inflation introduces information where none existed, no "best form" is possible. By limiting the candidate of choices to a very small set, final determination is simplified. The geometric models which frame the problem under study allow analysis either by interactive or by automated means. Both were described.

Analytical inflation is defined without recourse to a specified source of illumination. Instead, maximization of spectral invariance with respect to change in illumination is optimized. As with the the additive methods of *gamut mapping,* a final selection may then be made based on problem specifics.

With tools for both deflation and inflation in place, the analysis of spectral data comes full circle. Printer gamuts are sampled empirically, producing mosaics, thence abstract linear basis vectors which support the definition of reflectance solids

[12] Points appear out of gamut when in fact the centre of each dot defines its chromaticity.

having high symmetry. These in turn allow the creation of bases of reduced or increased dimension, thereby supporting the mapping of reflectance coordinates, ultimately describing once again the surface colours of naturally occurring objects.

# Chapter 8

# Extensions and Applications

The final chapter outlines further extensions to the thesis, with possible applications discussed in parallel. Four subsections follow which describe the partial exploration of extensions and related topics. These are arranged in order of depth of exploration. In §8.1, a complete data reduction is produced from a mosaic of 4096 tiles, the largest produced to date. The data reduction was undertaken using commercially available packages; representative command scripts are included for those who may follow. In §8.2, reflectance coordinates are represented in discrete pixels, allowing for direct spatial display of such data. A paper design of the system architecture is presented. In §8.3 the conditions of orthogonal are first extended to a wider class of basis functions, then weakened, allowing consideration of basis elements which are merely independent. The chapter concludes in §8.4 with a discussion of both colorimetric and spatial (rendering) models and their interplay.

## 8.1 Transmissive Film Characterization

Analysis of the Ektachrome 6184 sheet film[1] performed at the Canadian National Research Council required the custom application of techniques for data reduction and exchange, outlined in the first appendix of (Paeth, 1989a). The enormity of the $16 \times 16 \times 16$ data set prevented its full reduction during the thesis's research. However, the spatial permutation of tiles allowed the determination of bases from random samplings of the printer gamut by retaining as many of the initial mosaic elements as could be accommodated. Visual comparison of results were encouraging – a sampling of a few hundred points produced basis elements later seen to approximate a full reduction.

The arrival of custom linear algebraic software near the conclusion of the write-up provided a full decomposition. Nonetheless, computational overhead remains high: the full singular value decomposition (that is, finding 512 column vectors of length 512 from a $512 \times 4096$ data matrix) required 1.624 Gflops[2].

The final conversion required $372m16s$ computation time on a DEC5000/240 under Ultrix 3.5f[3]. Initial runs indicated that a substantial fraction of the running time was dedicated to swapping on a system having a 96 megabyte resident memory. A custom kernal with increased swapping space solved the problem; the final run required $232.4M$ of virtual memory, a $3 : 1$ ratio of virtual to physical memory required that the software job preempt all other timesharing tasks, thereby maximizing available resident memory.

[1]Thanks are due Murry Oles of Batten Graphics, Toronto, for providing the CT2T tape to film conversion service.

[2]A "gigaflop" represents $10^9$ floating point multiplications

[3]Creation of a $4096 \times 4096$ covariance required $289m\ 41s$.

### 8.1.1 Data Reduction

A licensed version of the extended MATLAB was instrumental in operating upon matrices of this size. The language was largely upward-compatible in comparison to the public domain version used to verify and test the SVD methods described in §5.1.4 and §5.1.5. The scripts were written to take advantage of a slightly modified syntax. More important, all sample and reduced data data matrices were converted into (new) MATLAB image format; additional matrices of reduced size use a companion ASCII representation in scientific notation.

The facility of the extended version is demonstrated in the following script. The disk file `svdevec.mat` contains $m$ SVD-derived transmission[4] vectors presented as a tabular text file of $n \times m$ entries.

```
load svdevec;
svd3 = svdevec(:,1:3);

dim = size(svd3);
rows = dim(1,1);

whiteref = ones(rows,1);

whitecrd = svd3' * whiteref;
whiteapr = svd3  * whitecrd;

hi = max(whiteapr);
lo = min(whiteapr);

whiteval = whiteapr/hi;

save whiteapr.txt whiteapr -ascii
```

[4]Throughout this section "transmission" substitutes for "reflection".

```
save whitehi.txt hi -ascii
save whitehi.txt lo -ascii
save whiteval.txt whiteval -ascii
```

The script computes the transmission coordinates `whitecrd` of an $n \times 1$ a white transmission `whiteref` on the 3D SVD subspace defined by `svd3`. The approximating transmission `whiteapr` is then computed and its minimum and maximum values recorded. This script was used verbatim in the production of Figure 4.5.

This commercial software supersedes a number of custom tools, including FIGBASIS, MATMUL and SPECINTERP. In their place the tool CIEINTERP was created to map a list of wavelength values (a column vector of sampling dimension $n$) onto a $n \times 3$ matrix of tristimulus vectors. By this a discrete representation of the tristimulus functions is created whose length conforms to the other data matrices. This allows straightforward conversions into XYZ or transmission coordinates using simple matrix multiplications; the creation of $(x', y')$ chromaticities under arbitrary source are also straightforward. A sample fragment of the scripts used to produce the figures in this section appears below, in which `s`, `svdevec`, `mssmevec` and `tristim.txt` respectively are the $512 \times 4096$ sample matrices, the two $512 \times 16$ basis matrices and the external text (matrix) as created by CIEINTERP:

```
load s;
load svdevec;
load mssmevec;
load tristim.txt;

% sample tristimulus values
xyz = s'*tristim;
save xyz.txt    xyz    -ascii

% sample chromaticity coordinates
```

```
    n = xyz * [1 1 1]';
   xy = [ xyz(:,1) ./ n, xyz(:,2) ./n ];
  save xy.txt     xy     -ascii

% sample svd transmission coordinates
  svd  = s'*svdevec;
  save svd.txt    svd    -ascii

% sample mssm transmission coordinates
  mssm = s'*mssmevec;
  save mssm.txt   mssm   -ascii

% svd basis chromaticities

  sbeta = svdevec'*tristim;
      n = sbeta * [1 1 1]';
  sbetaxy = [ sbeta(:,1) ./ n, sbeta(:,2) ./n ];
  save sbetaxy.txt sbetaxy -ascii

% mssm basis chromaticities

  mbeta = mssmevec'*tristim;
      n = mbeta * [1 1 1]';
  mbetaxy = [ mbeta(:,1) ./ n, mbeta(:,2) ./n ];
  save mbetaxy.txt mbetaxy -ascii
%
  quit
```

### 8.1.2 Illumination Coordinates

A set of 4096 points in tristimulus space are created using the above script. These describe the additive gamut under illumination by source E. Plotting against the Y (CIE luminous efficiency) coordinate produces two plots identical in production to Figure 6.5 (compare at 64 points) except the print gamut here is for transmissive media and shows reduced non-linear effects.

Figure 8.1: Ektachrome CIE Y vs X ($16 \times 16 \times 16$ Gamut)

The Y *versus* X plot appearing in Figure 8.1 closely resembles a similar plot appearing as Figure 5.2 in (Stone, Cowan and Beatty, 1988). A Y *versus* Z plot appears in Figure 8.2.

The gamut's CIE chromaticity coordinates may be found under CIE source E using linear algebra and a scalar normalization step, also appearing in the previous MATLAB script. The gamut is plotted in Figure 8.3.

### 8.1.3 Transmission Elements

The transmission basis elements are now presented. Figure 8.4 plots the first three vectors created through application of unmodified SVD.

Note that the spectral domain is unusually large (extending well into the near *infra-red);* a method of domain reduction is discussed in the last subsection concluding §8.3.1.

The data is replotted using the MSSM-derived basis elements in Figure 8.5. Notice the similarity of MSSM element 3 to SVD element 2. At first glace this would suggest that the introduction of $\bar{\beta}_1(\lambda)$ has displaced the MSSM basis elements, implying a MSSM space of dimension $m+1$ to rival an SVD space of dimension $m$. Closer examination shows that the elements 2 and 3 are swapped in reference to the counterparts they most resemble. This change in ordering comes about because the second MSSM element better approximates the first SVD element, whose goodness of fit is lost subject to the MSSM symmetry constraints.

In fact, the element sets span nearly the same space, made clear by studying the rotation basis (after Tables 5.2). Appearing in Table 8.1, the permutation of the second and third elements is made clear, as are the nearly common spans of both space by dimension $m = 3$.

Figure 8.2: Ektachrome CIE Y vs Z

Figure 8.3: Ektachrome Gamut Chromaticity

Figure 8.4: Ektachrome SVD Elements 1-3

Figure 8.5: Ektachrome MSSM Elements 1-3

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0.935 | -0.077 | -0.344 | -0.034 | 0.020 | 0.003 |
| 2 | 0.355 | 0.223 | 0.903 | 0.070 | -0.037 | -0.006 |
| 3 | -0.007 | 0.972 | -0.237 | -0.006 | 0.002 | 0.000 |
| 4 | 0.007 | -0.013 | -0.078 | 0.997 | -0.009 | -0.001 |
| 5 | -0.006 | 0.008 | 0.045 | 0.014 | 0.995 | 0.017 |
| 6 | -0.001 | 0.002 | 0.008 | 0.002 | -0.021 | 0.999 |

Table 8.1: Basis Rotation: MSSM versus SVD (Transmission Data)

### 8.1.4 Transmission Coordinates and Gamuts

It is instructive to study scatter plots of the transmission coordinates of the gamut samples after the previous manner. In Figure 8.6 the coordinates $(x_1, x_2)$ of the transmissivities in the SVD space are plotted in the $(x, y)$ plane of the paper. (That is, an orthographic projection of the gamut onto the $z$ plane is made, inviting comparison with the enclosing transmission solid.)

In Figure 8.7 identical SVD plot is presented after substituting the axes $(x_1, x_3)$.

The SVD-style plots show some bunching near the toe (which indicates that near-black points, though having little transmissivity recorded by $x_1$, take on a characteristic hue well-represented by a positive contribution of $x_2$). A second, opposing branch is also obvious, forming a region of reduced density near the centre of mass. On closer examination, a set of lines having northwest to southeast orientation are visible between the outer boundaries of the figure. These are in fact successive planes whose printer "blue" component increases through sixteen discrete steps in the range $[0 \ldots 255]$. This is make especially during interactive plotting of

Figure 8.6: SVD Transmission Gamut: $x_1$ versus $x_2$

Figure 8.7: SVD Transmission Gamut: $x_1$ versus $x_3$

the data (replacing the plotting symbol after each block of 256 entries would have a similar effect). This also reveals that an increase in the "green" component leads to a point migration down and two the right along any line.

The data is now replotted in Figures 8.8 and 8.9, substituting the MSSM-derived basis elements. Since these span nearly the same space, the effect is to reorient the gamut by rigid rotation. Nonetheless, the plots show a better balance of points, less non-convexity and most important, has an increased area (despite a reduced scale along the $y$ axis for coordinates $x_2$). It might be easily fit by a bounding quadrilateral. This suggests a strong similarity between the sheaf parallelogons of cross-section (abstract reflectance solids) and empirically-derived reflectance coordinates (empirical gamuts). Suggested further study would extend the interactive modeler so that both the solids and their defining device gamuts could be plotted simultaneously.

## 8.2 Specialized Hardware

The following section describes display subsystems which treat pixels in more rigorous terms: variable-length descriptors recording either illumination or reflectance coordinates on a well-defined underlying basis.

### 8.2.1 An Illumination Coordinates Frame Buffer

Basic frame buffers employ RGB phosphors, as these maximize the visible gamut. (See in particular both §3.2.4 and §3.2.5.) Although the phosphor chromaticities underlying the traditional (NTSC) broadcast standards and the emerging HDTV model are well defined, frame buffers rarely provide hardware which accounts for

Figure 8.8: MSSM Transmission Gamut: $x_1$ versus $x_2$

Figure 8.9: MSSM Transmission Gamut: $x_1$ versus $x_3$

the true chromaticities of the illumination coordinates present in any pixels. As seen in [3.4] for pixels and display of known photometry the correction is a $3 \times 3$ linear change of basis.

Hardware implementations are not unknown: all commercial television sets must provide an orthogonal rotation to map $YIQ$ broadcast coordinates onto an $RGB$ basis which ultimately controls the voltage on the grid of the cathode ray tube. In this setting the basis of both the input and the phosphor is known. The traditional analog implementation forms a set of nine currents in reference to input voltages using a resistor matrix. These serve as the elements in [3.4]. Since currents sum when flowing into a common node (Kirchoff's law), a wired-OR provides the desired control signal for each output. Were the monitor or input signal subject to occasional change, nine potentiometers (mechanically controlled variable resistors) could in theory be substituted to provide an adjustable change of basis with no change in throughput.

Digital broadcasting and related technology suggest a model which provides the change of basis by performing a discrete matrix multiply. (By spatial analogy, Silicon Graphics Inc. was founded upon a frame buffer architecture which provided a $4 \times 4$ hardware multiply in the spatial rendering chain, thereby allowing for a real-time a change of basis. This analogy is important and is explored more deeply in §8.4.) Although time between successive pixels is 14 $nS$ under the Nyquist limit and in many implementations twice this (i.e., four times the frequency of the colour subcarrier at 3.58MHz), the change of basis is constant. In this case table lookup may be used to remap the input stream to a weighted output stream, which need only be summed.

A device-independent (illumination) frame buffer architecture regards pixels as CIE $XYZ$ descriptors and provides a table of presets to account for the specific

monitor phosphors. Accordingly, both the exchange and renderer output formats would can take advantage of this radiant data model. However, the added expense is largely unjustified as a $3\times3$ multiplication can in all cases reconvert data; most users of CRT displays are unaware that phosphor chromaticities do vary (see Appendix F) yet monitor calibration is not an impractical task (Cowan, 1983), (Cowan, 1986).

Low-level display hardware employing illumination coordinates other than RGB pixel descriptors are described (Buchanan and Pendergrass, 1980) using the opponent colour system proposed for this purpose (Faugeras, 1979). In addition, researchers at Xerox (Buckley, Green et al., 1987) have proposed a frame buffer for a proprietary workstation which records data in the L*u*v* system. This has practical virtue, but places the burden of non-linear transformation upon the hardware in order to produce RGB values or their XYZ tristimulus equivalents.

### 8.2.2 A Reflectance Coordinates Frame Buffer

The direct display of reflectance data has been proposed (Taylor, 1973). Remarkably, the (remote sensing) data was mapped into 3-space using the SVD (Taylor, 1974b). Preliminary hardware table-lookup methods for reflectance-based data also known (Bass, 1981). The linear transformation of reflectance coordinate data is well-supported by an abstract pixel model (Fiume, 1986), (Fiume, 1989), especially one allowing an arbitrary[5] number of fields, wherein each component carries a predetermined (bit) precision (Paeth, 1986). Mapping a $j$ to a $k$-component pixel by $j \times k$ multiplication employing a matrix of constant elements is embodied by the tool IMMAT. The author's design takes heavy advantage of lookup tables of modest length in order to maximize throughput. A hardware implementation is a direct

[5]Of particular note pixels encoding more than three colour components

extension of the software. Its design became the primary vehicle of application in conjunction with the subtractive models proposed during this thesis's second stage examinations (Paeth, 1987b). The present discussion is derived from a paper-design completed by the author in 1990.

### Abstract Pixel Models (Background)

A frame buffer's memory need not store the literal RGB drive values presented to the CRT circuitry. In the traditional frame buffer, colour look-up tables (LUTs) are used to linearize the outputs of the RGB phosphors (Catmull, 1979). Since their creation, LUTs have been used to perform general effects (Blinn, 1979) such colour table rotation (Shoup, 1979), (Booth and MacKay, 1982). More abstract pixel models, whether implemented in hardware or software, are generally unexplored. Specifics of the bit versus integer-index model used here are discussed elsewhere (Paeth, 1990b).

A LUT can encode abstract information other than pixel intensity. An early extension is the $Z$ (depth) information used to describe distance to the surface of the object whose colour is recorded at that pixel (Gordon and Reynolds, 1985). More generally, alpha channels (Duff, 1985), subpixel coverage (Peachey, 1987) or surface orientation (Pique, 1983) may be recorded. The last example describes a system which allows for real-time illumination of scenes by recomputing tables based on the changing location of a hypothetical source. Because most LUTs provide an input index having only eight to ten bits, that implementation is restricted merely to changes in overall image intensity for test source located on a coarse angular grid.

Hardware systems in the early 1980's introduced a cross-bar as suggested by

Fuchs (Booth, Forsey and Paeth, 1986) allowing an arbitrary "swizzle" of bits in the integer channels which comprise a pixel. In more modern systems, the LUTs and DACs (digital to analog converters) are typically integrated, given the demands of high video rates and specialized circuitry. Although this limits the degree of experimentation within the traditional video chain, many modern frame buffers provide microcode-based local computing with speeds sufficient to remap pixels under simple transformations at video rates, rendering most hardware crossbar switching obsolete (Whitten and England, 1987).

A more abstract view of "uncommitted" pixel memory appears early in the creation of the Pixel-planes model (Fuchs et al., 1982), whose first functional architecture was designed by this author (Poulton, Fuchs and Paeth, 1985). Software-based extended pixel memory appears in the "Palette" system created at Waterloo (Higgins and Booth, 1986) which models paint as a semi-opaque (reflective) media and not under the traditional mixing of light. The latter system employed seventy-two bit virtual "pixels" coupled with an outboard physical frame buffer employing either a $512 \times 512 \times 24$ or a $1024 \times 1024 \times 8$ memory format. Its implementation took immediate advantage of the IM file format (Paeth, 1987a) for image storage and exchange.

### Reflectance Pixels

Pixels may be created such that each component records a reflectance coordinate on a suitably defined reflectance space. The basis vectors which define the space can change on a per image basis, but are constant during the RGB display of the image. At present, reflectance coordinate data arranged as a 2D raster is not common. For the purposes of study, the pixel output of the solid modeler described in Chapter 7 may be used. Alternately, a traditional scanner or camera equipped with filters

may be used in a manner analogous to the mask-making steps used by a commercial printer to produce colour separations.

Using the models described in Chapter 5, the reflectance coordinates define pixels whose components (other than the first) are signed values. This is a departure from most other common colorimetric models which employ positive components, most often on the range $[0 \ldots 1]$. This poses little problem, as the bit patterns of components are immaterial when used as indices. When occurring within table entries, two's complement arithmetic allows both summing and differencing to take place, that is, the implicit transformation matrices may contain negative elements.

The components can well take advantage of variable bit precisions. This is suggested by the decrease in sensitivity as indicated by the eigenvalues derived by the SVD. Thus, careful attention must be paid equally both to the number of components retained (dimension of the reflectance space) and the discrete precision of each component[6]. These suggest an initial reflectance component of maximum (eight to ten) bit precision, with additional components of reduced precision, until the pixel is filled. (For purposes of study, a thirty-two bit machine integer is a representative length.) At lower bit precision, trailing components do well to depart from the standard "bit-planes" model. Besides allowing greater selection of discrete levels, this allows the possibility of an odd number of levels, of importance where the asymmetry of two's complement representation dominates. As an example, a four-bit descriptor apportioned into $(2, 1, 1)$ bit planes provides one channel with values $[-2, -1, 0, 1]$ and two channels with values $[-1, 0]$. By comparison, by regarding the pixel as a Cartestian product of channels[7] (Paeth, 1990b), three equivalent channels with symmetric bipolar ranges $[-1, 0, 1]$ are possible. A set of factorizations useful

[6]Thanks are due to John Keightley of Pthalo Systems for this insight.

[7]That is, components neither require $2^m$ states nor must be aligned to bit boundaries

| N | Factors | $\log_2 N$ |
|---|---|---|
| 15 | $3 \cdot 5$ | 3.91 |
| 27 | $3^3$ | 4.75 |
| 63 | $7 \cdot 9$ | 5.98 |
| 121 | $11^2$ | 6.92 |
| 125 | $5^3$ | 6.979 |
| 225 | $3^2 \cdot 5^2$ | 7.81 |
| 231 | $3 \cdot 7 \cdot 11$ | 7.85 |
| 243 | $3^5$ | 7.92 |
| 245 | $5 \cdot 7^2$ | 7.92 |
| 247 | $13 \cdot 19$ | 7.92 |
| 255 | $3 \cdot 5 \cdot 17$ | 7.99 |

Table 8.2: Useful Pixel Factorizations (Odd States)

with eight-bit data is presented in Table 8.2, in which the third column estimates total (binary) bit precision.

### Bit Permutation

A bit permuter allows the creation of a frame buffer in which the number of channels and the bit precisions of each can be changed arbitrarily. Table lookup provides the hardware designer's method of choice in forming a generalized cross-bar switch used to reroute bits (Lawrie, 1975) in channels of moderate width[8]. Further, the tables need not provide a mere shift or scatter of their input values, but can additionally provide an arithmetic map. For instance, a memory table containing 256 (32-bit)

[8]In larger systems, networks implementing the *perfect shuffle* are employed (Batcher, 1968).

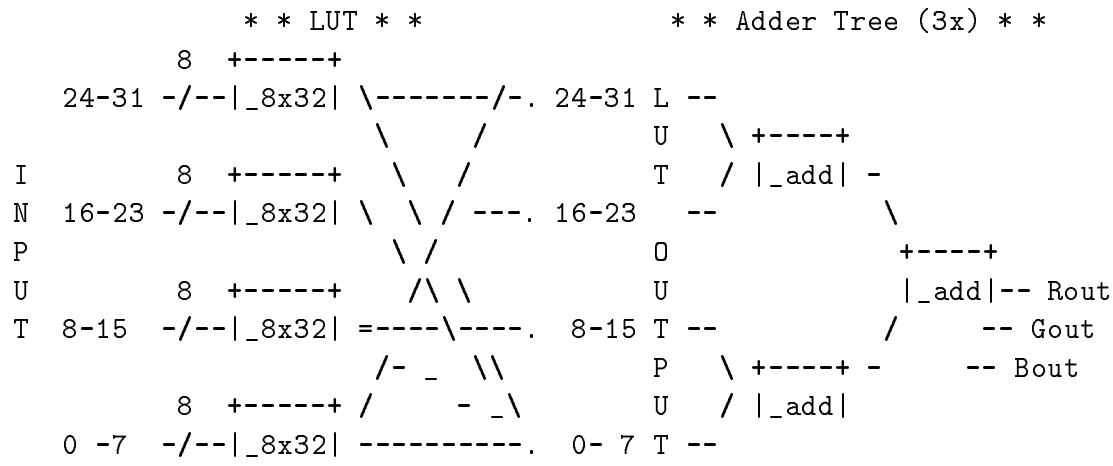



Figure 8.10: High-Performance Pixel Datapath Mapper

long words provides an arbitrary mapping from eight bits onto thirty-two, yielding a fully general LUT for a system having an eight-bit data path from the pixel memory. A 32-to-32 mapping may be constructed elegantly by using wire ORing (ANDing), as with open collector logic, with unused bits are left at zero (one, by DeMorgan's theorem). A sketch appears in Figure 8.10.

Here the 32-bit pixel data enters from the left and flows through the parallel LUTs. Next, the $4 \times 4$ bipartite graph represents interconnect from the outputs of each $8 \times 32$ memory to the adder tree. (That is, each intermediate output at position $O$ is the direct wiring of four memory outputs at positions $O$, $O+8$, $O+16$ and $O+24$; each group of eight quadruples is represented by the '.' which indicates the point of common interconnection.) Such a device allows an arbitrary mapping of thirty-two bits onto thirty-two bits, given suitable table entries.

Constructing such as device as a replacement cross-bar is valuable. The original

Ikonas cross-bar employs instead thirty-two discrete multiplexers, allowing any output to be drawn from from any of 32 (in fact, 34) input sources. Not only must the design employ 32 24-pin MSI packages[9], it is less general. This static RAM design presented here is considerably more compact, uses standard sized packaging, and is more general.

The right-hand block in the diagram is the adder tree. Its sums the intermediate output terms to create what are the row sums of the constant $3 \times 4$ weight matrix, thus creating the final R, G and B pixel outputs. The operation is strictly combinatorial and hence unclocked, save for the pipeline delay needed to allow all summations (and LUT routing propagations) to complete. This is accommodated by fetching the final output a discrete number of pixel clocks later – the overall effect merely shifts the image on the display. This simplicity of design makes it directly amenable to use on a frame buffer such as the Ikonas in which the crossbar board is directly interposed in the video chain – in fact, has matching input and output cable headers to allow its physical and electrical inclusion or exclusion. Conversely, it must run at video rates (or provide a pipelined output at some integral pixel delay). Fortunately, it is the random-access memory (RAM) chips which receive the chip designer's utmost attention in minimizing propagation delay.

The added generality of the wird-OR LUT approach allows blocks of eight bit inputs to be remapped (as indices) before scattering. For instance, $\mu$-law encoding of the input which reduces quantization effects across the full dynamic range (see (Paeth, 1990c) and the related colour plates appearing in *Transactions on Graphics*) can map an eight-bit input descriptor into eight bits of output, yet present parallel results in the form of three copies at locations 0-7, 8-15, and 16-23, allowing a monochromatic RGB rendering from a single input channel. More general effects,

[9]These are 16-to-1 demultiplexor/data selectors.

such as chroma keying, have been proposed by the author and are the subject of another paper.

### The Mathematical Model

Through use of the table lookup and subsequent address tree, pixel data is not merely merged but summed together to form linear R, G and B channels. That is, any mapping of the form

$$\begin{aligned} R_{out} &= T1[R_{in}] + T2[G_{in}] + T3[B_{in}] + T4[O_{in}] \\ G_{out} &= T1[R_{in}] + T2[G_{in}] + T3[B_{in}] + T4[O_{in}] \\ B_{out} &= T1[R_{in}] + T2[G_{in}] + T3[B_{in}] + T4[O_{in}] \\ O_{out} &= T1[R_{in}] + T2[G_{in}] + T3[B_{in}] + T4[O_{in}] \end{aligned} \tag{8.1}$$

can be constructed in parallel. When the byte table entries define a linear scaling of their index, they serve as a hardware multiply, by analogy to the resistor/current-summing model described previously. In this case, a real-time $4 \times 4$ matrix multiplication takes place within the video chain, thus providing the linear change of basis. This simplifies [8.1] yielding

$$\begin{aligned} R_{out} &= W_{1,1}R_{in} + W_{2,1}G_{in} + W_{3,1}B_{in} + W_{4,1}O_{in} \\ G_{out} &= W_{2,1}R_{in} + W_{2,2}G_{in} + W_{3,2}B_{in} + W_{4,2}O_{in} \\ B_{out} &= W_{3,1}R_{in} + W_{2,3}G_{in} + W_{3,3}B_{in} + W_{4,3}O_{in} \\ O_{out} &= W_{4,1}R_{in} + W_{2,4}G_{in} + W_{3,4}B_{in} + W_{4,4}O_{in}. \end{aligned} \tag{8.2}$$

(The "O" overlay channel is provided for hardware compatibility; only the R, G and B values are required in creating an R-buffer). In practice, the weights $W$ are computed given the spectral power distribution of some hypothetical source. These

may be adjusted after the contents of the frame buffer memory is loaded, allowing a change in source of illumination without need to recompute a pre-rendered scene (see also §8.4).

The basic eight-bit (times four) model is presented here because its implementation is particularly straightforward using conventional components. A more advanced model might apportion a thirty-two bit reflectance pixel into channels having $(9, 7, 6, 3, 3, 2, 2)$ bits (this by inspection of the base two logarithms of the SVD weights computed previously); an empirically-derived would take advantage of a non-Cartesian representation scheme. This practical model employs a wider (twelve-bit) data path with extended tables to support the greater ($> 8$) bit pixel precisions of the $x_1$ channel that may be encountered in practice.

## 8.3 Extended Models and Related Applications

The reflectance bases which span reflectance spaces are subject to the addition or deletion of defining constraints. These changes widen the scope of application of the methods set forth at a potential sacrifice of properties previously described.

### 8.3.1 Extended Orthogonal Bases

#### Harmonic Functions

The initial elements defining an orthogonal reflectance basis may be further predefined, subject to overriding proviso that the first element be constant. For instance,

the set of periodic functions

$$\begin{aligned} \beta_1(\lambda) &= 1 \\ \beta_2(\lambda) &= \cos\lambda \\ \beta_3(\lambda) &= \cos 2\lambda \end{aligned} \qquad [8.1]$$

form an orthogonal set in which the first element of the series is unity. Additional basis vectors may be found using the MSSM method in which all three orthogonal components are removed from the covariance matrix before the eigensystem is solved iteratively.

The advantages of such a system are that the reflectance coordinates derived under dissimilar empirical data nontheless share the first three basis elements, allowing the direct exchange of data. Put another way, when any two empirically determined reflectance spaces $\mathcal{R}^m$ and $\mathcal{R}^{m'}$ are deflated onto $\mathcal{R}^3$ they have an identical reflectance solid. At first glance the harmonic functions of [8.1] would suggest spherical surfaces having multiple nodes, but in fact the convexity of the reflectance solid cannot allow for this. (Plots of these functions is a straightforward task but was not done.)

The disadvantage of this system are two-fold. First, there is additional loss in goodness of fit by presupposing a general *a priori* solution in the absence of empirical data – harmonic analysis is a mathematical device in no way suggestive of typical reflectance functions. Second, the ordering of basis elements by goodness of fit no longer holds. Again, a fourth basis element might provide better fit then elements two or three. Still, a canonical ordering may be established by defining a nested sequence of reflectance spaces in which the first basis elements of the predetermined sequence taken in order, followed by the remainder determined by empirical means.

**Tine Functions**

A regrouping of the tine elements described in conjunction with Figure 2.2 allow orthogonality by virtual of the non-overlap of the functions anywhere within the domain of wavelength. As above, three (e.g.) basis elements are defined as canonical without regard to the variance they minimize. Trial elements are:

$$\begin{aligned} \beta_1(\lambda) &= 1 \text{ iff } 557.7\text{nm} \leq \lambda < 700.0\text{nm} \text{ else } 0 \\ \beta_2(\lambda) &= 1 \text{ iff } 490.9\text{nm} \leq \lambda < 557.7\text{nm} \text{ else } 0 \\ \beta_3(\lambda) &= 1 \text{ iff } 400.0\text{nm} \leq \lambda < 490.9\text{nm} \text{ else } 0 \end{aligned} \qquad [8.2]$$

which partition the interval of visible wavelength [390nm ... 710nm] at two points. This scheme has the further advantage of providing a general estimate for the red, green and blue appearance when the surface is illuminated under a broad-band, achromatic source (Barlow, 1982). Consequently, the reflectance coordinates produced may be rendered directly on a conventional (illumination) frame buffer; omitting the fourth and higher-order coordinates constitutes a deflation. The 3D reflectance solid common to any space derived with this partially predetermined basis is the cube. The method may be extended to higher dimensions by grouping (summing) tine elements into distinct sets, thereby forming comb reflectance functions which are piecewise constant. The related solids are all (hyper)cubes, formed by rigid rotation of the parent cube by ninety degrees. Enforcing non-negativity on the basis vectors has the useful property of guaranteeing non-negative reflectance coordinates. This is analogous to the CIE's choice in deriving a non-negative basis set (the tristimulus vectors) from a set of colour matching functions which exhibit negative regions.

The partitioning wavelengths defining the trial basis in [8.2] above are the inner two sampling wavelengths taken from the set {456.4, 490.9, 557.7, 631.4} derived experimentally (Meyer, 1986), who advocates the adoption of reflective

colour spaces in computer graphics (Meyer, 1988). In fact, his results define an orthogonal tine model, thereby enjoying many of the benefits previously set forth. However, its drawbacks include chromatic aliasing (an abrupt jump in hue, as when imaging a rainbow) plus the lack of an equipotential reflectance. A generalization of this extended tine basis (whose elements have values on the set $\{0, 1\}$) considers those elements which take on values $\{-1, 0, 1\}$. This approach is unexplored, but quickly leads to a study of the Walsh/Hadamard functions and the related theory of wavelets. (See also §4.1.10.)

#### Interval Tapering

The data reductions performed with this thesis applied regression techniques upon spectral reflectance vectors sampled across an extended visible domain. While this spectral oversampling is often advocated for practical reasons, there are drawbacks. As an example, most colour transparency stock is transparent in the *infra-red,* as significant absorption of heat rays would accelerate the deterioration of projected images. Regression techniques can potentially capture the profile of reflectances (transmissivities) in this region, leaving a set of model vectors which minimize variance for wavelengths which are largely invisible. Rather than limit the range of visible wavelength abruptly, a tapering of the visible interval can be produced to account for the decreasing sensitivity of the visual system for extreme wavelengths. For the sake of simplicity, this was not done.

### 8.3.2 Non-Orthogonal Bases

Reflectance bases whose elements are merely independent lose many of the benefits previously enjoyed, yet are still valuable. As an example, an inflation $\mathcal{R}^2 \xrightarrow{\mathrm{I}} \mathcal{R}^3$

reveals two axes of representation as when mapping reflectance data having two components onto a colour printer. Given the constrained basis element $\bar{\beta}_1(\lambda)$ and a second element $\beta_2(\lambda)$, derived through inflation, the adjusted element

$$\tilde{\beta}_2(\lambda) = \bar{\beta}_1(\lambda) - \beta_2(\lambda) = 1 - \beta_2(\lambda) \qquad [8.3]$$

formed through linear combination defines the same space under the change of basis. There is no representational loss in the system: goodness of fit as determined by methods of orthogonality is unchanged. Geometrically, the reflectance solid remains unchanged but is rotated into a symmetric position such that its grey axis lies in the plane (and is the average of) the two independent basis vectors.

Since both elements sum to one, all "grey" reflectances $\rho(\lambda) = c$ have reflectance coordinates $x_1 = x_2 = c$. This suggests the partitioning of the spectral domain in the previous section, except both reflectances are now continuous and multivalued across the entire domain. Linearity assures that the reflectance coordinates of linear combinations of reflectance functions are the combinations of the coordinates themselves. Thus, a complementary colour $(x, y)$ may be easily determined by subtracting it from white $(1, 1)$; the complement lies at $(1-x, 1-y)$. Simple algebra then shows that all saturated colours have coordinates $(x, 1 - x)$; transposition $(1 - x, x)$ finds the saturated opponent colour.

A set of representative cross-sections was created using the interactive modeler by applying the sheaf methods introduced in §7.2.3. The reflectance function associated with each chosen belt vertex is plotted together with its conjugate. Three representative angles maximizing belt girth were chosen (see Figure 7.12): these vertices are persistent features, after §4.2.10. The plots appear in Figure 8.11.

The functions are the *opponent reflectances* described in both §5.2.2 and §7.3.3. The complementary reflectance appearing thus belongs to the opposing vertex,

Figure 8.11: Complementary Reflectance Functions (3)

whose conjugate angle[10] appears after the slash. The functions (solid) show an evolution from low-pass through band-pass; their conjugates accordingly model high-pass and notch (band-stop) filters.

Used as basis elements, the functions create an independent coordinate system which maximizes dynamic range of the reproduction, provides a neutral grey given coordinate having similar components and and highly saturated hues given dissimilar components. This scenario is strongly analogous to the *gamut mapping* of trichromatic coordinates lying in 3-dimensional illumination space (Stone and Wallace, 1991). A feature of that work are the stretches along the grey axis or small rotations (so as to minimize the introduction of colour shifts) used to fit input device gamut to output device gamut, all within the additive domain, while maintaining high colour fidelity. Here, the fitting is not done with regards to any source, though final choice of inflation axis may be driven by thematic use or to otherwise choose a representative hue associated with the new axes.

As an example, images of Hamilton, Ontario sampled on September 20, 1985 were acquired from a Landsat 5 Thematic Mapper[11]. Its sensors are optimized not for the visual system (RGB primaries) but instead for studying the reflectivities of chlorophyll (Li and Strahler, 1985). Its second channel ($520 - 600$nm) was mapped onto reflectance coordinates using the function appearing in Figure 8.10(c), thus maximizing dynamic range of the SVC6500 device gamut while preserving colour authenticity. The neutral axes were formed as a weighted sum of the first ($450 - 520$nm) and third ($630 - 690$nm) channels using conventional methods. This application is an empirically-based extension to the CRT-based *pseudo-colour*

[10]This term is used to describe an angle sum of 360°; conventional terms such as complementary (sums to 90°) and supplementary (sums to 180°) do not apply.

[11]Thanks are due to the Mapping, Analysis and Design Group of the Faculty of Environmental Studies, University of Wateroo.

methods common on CRT monitors (Booth and MacKay, 1982). Here, the plot takes on a green versus magenta hue of varying intensity. Initial image production (not appearing) encourages further research.

## 8.4 Spatial Extensions

Throughout this thesis, reflectance functions $\rho(\lambda)$ have been considered independent of spatial effects. Interestingly, the bidirectional reflectance (distribution) functions or *BRDF* which underpins much current work in radiosity considers spatial geometry carefully, but places less attention on colour representation. Clearly, an extension of this work is the integration of both emerging techniques.

### 8.4.1 Modeling Spatial Effects

State-of-the-art rendering systems necessarily model physical reality to a fine degree. Ultimately, they must deal with the interrelation of spatial and colorimetric effects. Early treatment of surface reflections in the context of computer graphics is due to (Blinn, 1977). Both diffuse and specular reflections can be treated by methods of linear interpolation (Whitted, 1980) (Hall and Greenberg, 1983). The "gloss" model of (specular) highlights has an underlying analytical model (Cook and Torrance, 1981) which may be derived from physical study (Torrance and Sparrow, 1967). Other methods may also be classified as analytical (Blinn, 1982) or empirical methods (Cabral, Max and Springmeyer, 1987), (Immel, Cohen and Greenberg, 1986).

Ultimately, the physical reality of colour as a quantum mechanical effect appears when spatial feature size is on the order of the illuminant wavelength (Smits and

Meyer, 1990). Dichroic colour, as seen in the peacock plumage or in oil on water are common examples and show strong angular dependence. Likewise, scattering effects such as the Blue Jay's plumage manifest themselves. These and related effects (such as polarization) are not considered. Instead, extended reflectance functions are embodied exactly by the BRDF.

A BRDF which is constant models a surface having *Lambertian* reflectance: light is diffused equally in all directions; no spatial function is required. In this case no extension to the basic model $\rho(\lambda)$ is needed. At the other extreme, *specular* reflectance[12] as from metallic surfaces reflects a ray along a highly select path: the angle of incidence matching the angle of reflectance. Here it is the colorimetric function which is constant, as $\rho(\lambda) \approx 1$ defines the specular component of reflection.

Traditional rendering algorithms model specular effects to achieve a degree of realism. An early example is the simulation of highlights present in Phong shading (Phong, 1975) based on surface orientation. The technique persists as even advanced radiosity-based methods often employ a final Phong pass to provide realistic "sheen", as in the two-pass method (Sillion and Puech, 1989) based on the foundational radiosity rendering equation (Kajiya, 1986) (see also plate 40 in (Cohen and Wallace, 1993)). More recently, radiosity methods have encorporated models of directional reflectance, including the classification of transport paths (*op. sit.* §10.2) to account for rays representing general (diffuse) illumination or modeling a highlight path (Rushmeier and Torrance, 1990). Most often, these employ a reflectance model in which the BRDF is split into a Lambertian and a specular component (Heckbert, 1991). This makes the BRDF a prime candidate for representation using the symmetric reflectance bases described in Chapter 5.

[12] This is derived the Latin *speculum* for mirror, as in "speculum metal".

Simply put, *specular reflectance is modeled exactly in any reflectance space*, because in all cases the basis element $\bar{\beta}_1(\lambda) = c$ is present. That is, sheen may be introduced to a reflective surface merely by adjusting the reflectance coordinates $x_1$ of the surface upwards for cases when the angle of incidence approaches the angle of reflectances. Put another way, the specular component can be modeled exactly using a a symmetric reflectance basis having only one (constant) element.

Extensions of this method modeling BRDFs using an approach similar to the local bases examined in §6.5. Here, a global reflectance basis is used to characterize the surface generally, local adjusting bases then capture secondary effects of sheen or spatially variant colour. Empirical study would require a test assembly devoted to the task, as the BRDF must be sampled at many orientations to produce adequate data for the regression methods (Murray-Coleman and Smith, 1990). It is conjectured that the second order basis could be of greatly reduced dimension (e.g. two) in order to capture the spatial colour dependency as a second-order effect. This is suggested by empirical studies of BRDFs which when modeled using a series of spherical harmonics (analogous to the Fourier series) can be greatly truncated (Sillion et al., 1991) once the "ideal specular term" has been accounted for.

Regardless, a study is strongly suggested based upon set of reflective surfaces in which both multiple geometries and surface samples are varied. The basic study would assume an isotropic reflectance function (compare (Kajiya, 1985) and (Poulin and Fournier, 1990)) and thereby add merely one degree of freedom the angle formed between source and detector during sampling[13].

[13]All reflectance data characterized in the course of this research used a 45°/0° sampling geometry.

### 8.4.2 Integrated Colorimetric and Spatial Rendering

With a general BDRF in place, the rendering of a scene proceeds in two ways: both by position and by source of illumination. Traditionally, rendering describes merely the spatial task[14]. In many systems, a rudimentary colour descriptor (such as an RGB value, adjusted by brightness or under conditions of anti-aliasing) define the colour as presented on the display: no colorimetric rendering takes place. The very latest systems compute colour using restricted sampling (low dimensional tine) models and provide output in CIEXYZ or in $(Y, x', y')$ illumination coordinates, preserving carefully crafted realism across different output devices.

#### Permutation of Rendering Order

The equal status of spatial and colour rendering suggests the commutation of their order of execution. A alternative is now advanced: *the spatial rendering is conducted without any knowledge of the spectral power distribution of the sources in the scene.* The computation is in accordance with the reflectance composition equation [2.23]. The output of such a system is a set of reflectance coordinates which have been projected from a space of continuous, spatial "world" coordinates onto the discrete 2D coordinate system of the display image. The output can be immediately displayed on a reflectance-based frame buffer, or otherwise converted into RGB coordinates for some source. Although a basic implementation does not admit the possibility of multiple sources having known position but unknown spectral power distribution, it is nontheless well-suited to the modeling of natural, out-of-doors scenes in which the distribution of the source (sun, moon) changes with cloud-cover or time of day.

[14]This remains true even in many radiosity-based systems which are otherwise state-of-the-art, as for example (Buckalew and Fussell, 1989).

Here the methods of deflation are particularly well-suited to overhauling the most advanced radiosity software by supplanting the tine basis with one having better fit, thereby allowing either an increase in realism with constant data structure size or a decrease in computational time with constant accuracy.

### The Rendering Analogy

Traditionally, abstract objects in a scene are described in both spatial and colorimetric terms. Spatial rendering is the mapping from three-dimensional *world coordinates* onto a discrete set of two-dimensional *screen coordinates* in reference to a given eye point. By analogy, colour rendering maps continuous surface reflectances into chromaticity coordinates (or any independent triaxial system of perceptual colour, such as the linear RGB of phosphors) for a given source of illumination. The analogy is deep because both models exist in reference to the sensory apparatus of the human visual system: the fovea of the human retina is comprised of a a two-dimensional spatial array of cones; the cones fall into three detector classes. Thus, the lowest dimensional model which can provide an illusion of the world must provide two spatial and three colorimetric dimensions. The analogy is deeper still (suggesting the unified model described above), because both geometric projection and cone responses follow a linear model. These similarities are tabulated in Table 8.3.

In much the same way that 3D world coordinates of spatial rendering are often formulated (and sometimes implemented in hardware) using a 4D system of *homogeneous coordinates,* the *symmetric colour space* described throughout this thesis allow the representation of reflectances using linear models of high dimension, ultimately presented as low dimensional coordinates for purposes of visualization.

| Terminology | Spatial Model | Colorimetric Model |
|---|---|---|
| perceptual transform | $3D \rightarrow 2D$ | $nD \rightarrow mD \rightarrow 3D$ |
| point of projection | eye point | source of illumination |
| point of data | world coordinate | reflectance coordinate |
| hardware assistance | yes | proposed |
| enabling model | homogeneous coordinates | symmetric colour space |

Table 8.3: The Spatial/Colorimetric Rendering Analogy

### 8.4.3 Other Effects

Fluorescence can be handled directly in the computation of the total reflectance of a system of surfaces. After [2.23], multiplications are employed to model the subtractive effects. For a set of diagonal basis elements, the multiplication takes place on a per-wavelength basis, seen in §4.1.4. For general bases, a $m \times m$ matrix is used to create the change of basis which ultimately maps an input $m$-vector of reflectance coordinates to an output $m$-vector. (Note that no colorimetric properties of the eventual test ray are supposed, merely its eventual arrival. This is an instance of the permuted rendering order of the previous section. See also (Neuman and Neumann, 1989).) Adjustments to this array then account for the loss of energy at short wavelengths and their reintroduction at longer ones. The physical condition $\rho(\lambda) \leq 1$ need no longer hold, provided that energy is conserved. Further extensions to treat phosphorescence are not easily accommodated because the system transfer function must then save state.

Dissimilar bases may arise in practice should reflectance data on one bases require modeling or conversion on an alternate basis. The former situation suggests a basis of increased dimension $m + m'$ to represent data originating under either

basis. The latter suggests up to $m \times m'$ cross terms formed as the Cartesian product of the dissimilar sets of basis elements. In fact, the common element $\beta_1(\lambda)$ reduces significantly the total number of terms for small $m$.

# Appendix A

# Singular Value Decomposition

Successive solutions are possible using Weilandt deflation to remove a known eigenvector from the symmetric matrix using an outer-product form. Here $\mathbf{C}$ is the initial symmetric covariance matrix, $\mathbf{v}$ an eigenvector of $\mathbf{C}$, and $\mu$ its associated eigenvalue, such that $\mathbf{C}\mathbf{v} = \mu\mathbf{v}$. Then the residual matrix

$$\mathbf{C}' = \mathbf{C} - \mu\mathbf{v}\mathbf{v}^{\mathrm{T}} \qquad [A.1]$$

removes the eigenpair $(\mu, \mathbf{v})$ by application of the identity [4.15]. This is Hotelling's method (Hotelling, 1943). Iteration allows the removal of successive eigenpairs. The method is is generally stable (Paige, 1980), widely practiced (Watkins, 1991c) and serves as the basis for state-of-the-art block methods. The latter methods (Golub, Luk and Overton, 1981) which locate a set of eigenpairs *en masse* which are subsequently normalized using a QR (Gram-Schmidt) orthogonalization (Ruhe, 1983) decomposition prior to deflation to increase accuracy when solving full systems. The numerical accuracy of this and related iterative methods are described at length elsewhere (Stewart, 1973); a classic treatment appears in handbook form (Wilkinson and Reinsch, 1971).

The custom tool created here represents the limiting case with unit block size set to one. Only partial decompositions are required. A full decomposition is necessarily unstable since subtraction of like quantities manifests. Statistically, this indicates that the direction of the basis vectors is not well-defined when they must indicate an axis of least variance upon data which is uniformly distributed (in the noise). This is of no concern, as the deflation is terminated with the appearance of a near singular value for precisely the same reason: the related basis element provides very little additional representational accuracy.

The tools were verified against full decompositions created under MATLAB (to the latter's storage limit at $200 \times 200$). Diverge appeared only after the matrix was near zero. As an added precaution, double-precision arithmetic was substituted for the calculation of all intermediate quantities. Yet undescribed is the method for locating the largest eigenvector, which had arcane but unimportant consequences of algorithm design in the presence of certain colorimetric constraints.

A full C-language implementation appears below. The code also serves as an example of accessing the MATLIB procedures described in Appendix E.

```
#include <stdio.h>
#include <math.h>
#include "matlib.h"

#define ITERATE 100
#define EPS (1.0e-7)

/*
 * Hoetelling's "power" method for finding the largest eigenvalue and vector of
 * a Grammian matrix (symmetric, of the form aa', eg positive-semidefinite,
 * with all eigenvalues real and non-negative). The function returns the
 * largest eigenvalue and fills in the eigenvector matrix "v" with an eigen-
 * vector in each successive column. The "c" column vector is filled in with
 * the matching eigenvalue. The symmetric is NOT updated with its residual
 * as these are computed in double precision as required. Each eigenvector
```

```
 * is scaled as a unit vector.
 *
 * Note: this function is most often used for Singular Value Decomposition.
 * In this case, the sqrt() must be taken on each eigenvalue to form the
 * related singular value.
 *
 * Further note: in the presence of "flag", the first output vector produced
 * with be the "DC" vector of all ones; the reduction the proceeds from there
 */

/*
 * a - symmetric matrix of size [n..n]
 * v - eigenvector (columns) matrix of size [n..lim] (lim eigenvectors)
 * c - eigenvalue vector (row) of [1..lim] giving [e1 .. elim]
 * flag - true if v1 = [c c ... c] to be constrained
 */

mateigen(a, v, c, flag)
  mat a, v, c;
  {
  mat v1, v2;
  mat d, t;        /* for FIRST */
  double scale, oldscale, oldscale2, rsum;
  int it, i, j, e, r, first;
  int mi, mj, mk;

  matreg("eigenbig");
  matcheck(a);
  matcheck(v);
  matcheck(c);
  matconf("cols", C(a), R(v));
  matconf("cols", C(c), C(v));
  matconf("rows", R(c), 1);
  v1 = matalloc(C(a), 1);
  v2 = matalloc(C(a), 1);
  first = 0;
#ifdef FIRST
  d = matalloc(R(a), C(a));
  t = mattransa(v1);
#endif
  for (e=0; e<C(c); e++)           /* for each eigenvector to be found */
    {
    matconst(v1, sqrt(1.0 / (real)C(a)));  /* initial unit eigenvector */
    oldscale2 = oldscale = 1.0;
/*
 * form the successive products "cv' <- Av" where v' is unit length and c is
 * a constant (used to normalize its length) and repeat as "cv'' = Av'", etc.
```

```
 * until v' = v. Then by definition v is an eigenvector and c its eigenvalue.
 *
 * Factor out each successive eigenvalue as "A - cvv'" forming the residual
 * matrix (the technique of Wielandt Deflation). We form the RHS outer matrix
 * product in place to save the n^2 storage.
 *
 * In this code we compute the residuals from the partial eigenvector/value
 * table and make their effect implicit in the matrix multiplication. This both
 * increases precision and means that A is left uneffected. The expense is
 * making the unit cost eigenvector finder pay O(n^2) in finding the first n.
 */
#ifdef FIRST
    for(it = 0; it < ITERATE; it++)
      {                  /* converge on finding eigenvector */
      matmul(v2, a, v1);
      scale = veclen(v2);
      if ((scale == 0.0) || (flag && (e == 0))) break;
      matscale(v1, v2, 1.0/scale);
      if (e > 0)
        {
        rsum = 0.0;
        for (i=0; i<R(v1); i++) rsum += EL(v1,i,0);
        rsum /= R(v1);
        for (i=0; i<R(v1); i++) EL(v1,i,0) -= rsum;
        }
      }
    mattrans(t, v1);
    matmul(d, v1, t);
    matconst(d, scale);
    matsub(a, a, d);
#else
    for(it = 0; it < ITERATE; it++)
      {                  /* converge on finding eigenvector */
      for(mi=0; mi<R(a); mi++)
        {                /* matrix (residual) multiply */
        double sum, psum, nsum, res, pres, nres;
        for(mj=0; mj<C(v1); mj++)/* really just once! */
          {
          psum = nsum = 0.0;
          for(mk=0; mk<R(v1); mk++)
            {
            nres = pres = 0.0;
            for(r=0; r<e; r++)
              {
              res = EL(c,0,r) * EL(v,mi,r) * EL(v,mk,r);
              if (res > 0.0) pres += res; else nres += res;
              }
```

```
            res = pres + nres;
            sum = (EL(a,mi,mk) - res) * EL(v1,mk,mj);
            if (sum > 0.0) psum += sum; else nsum += sum;
            }
          EL(v2,mi,mj) = psum + nsum;
          }
        }
      scale = veclen(v2);
      if ((scale == 0.0) || (flag && (e == 0))) break;
      matscale(v1, v2, 1.0/scale);
/*
 * use a modified Gram-Schmidt to guarantee ortho(v1, v2..vn).
 * if "flag" is true then apply
 *
 *              (Vn . V1)
 *   Vn' = Vn - --------- V1
 *              (V1 . V1)
 *
 *       = Vn - (Vn . V1) V1           with V1.V1 = 1)
 *
 * so v1 = [c c ...c] = c [1 1 ... 1]  with c = 1/sqrt(n)  so
 *
 *       = Vn - (Vn . (c)[1 1 ... 1]) [c c ... c]
 *       = Vn - (c) (c) (Vn . [1 1 ... 1]) [1 1 ... 1]
 *
 * where c*c = 1/n and V . [1 1 1] = rowsum(V), we have
 *
 *   Vn' = Vn - (rowsum(Vn)/n [1 1 ... 1])
 */
      if (flag)
        {
        double rowsum = 0.0;
        for (i=0; i<R(v1); i++) rowsum += EL(v1,i,0);
        rowsum /= R(v1);
        for (i=0; i<R(v1); i++) EL(v1,i,0) -= rowsum;
        }
      if ((scale == oldscale) || (scale == oldscale2)) break;
   /* if (fabs(scale - oldscale) < fabs(scale*EPS)) break; */
      oldscale2 = oldscale;
      oldscale = scale;
      }
#endif
    if (it >= ITERATE)
      {
      fprintf(stderr, "\
WARNING: convergence on vec %d is only to within %g after %d\n",
           e+1, fabs(scale-oldscale), ITERATE);
```

```
   /* exit(1); */
      }
/*
 * update master matrices
 */
    for (i=0; i<R(v); i++) EL(v,i,e) = EL(v1,i,0);
    EL(c,0,e) = scale;
    }
  }
```

Thus, the column vector $[1, \ldots, 1]^{\mathrm{T}}$ serves as an initial guess. It is premultiplied repeatedly by the symmetric matrix until convergence to within `EPS` is achieved or until the limit `ITERATE` is reached. Matrix deflation is not performed, clear because the MATLIB matrix reference macro `EL(a,)` never appears on the left-hand side. Instead, an an *on-line algorithm* employed to compute adjustments as needed. Also note that residuals are tallied by sign in either `rsum` or `psum` and then summed once outside the loop. This reduces the summing of quantities of unlike sign to oneoperation, greating increasing floating point precision (Rice, 1983).

By happenstance, the modification which imparts symmetry to reflectance spaces is based on the same constant vector. Thus, preadjusting the data matrix (using the $MSSM$ method of §5.1.5) removes this axis (grey reflectance, body diagonal) from the covariance matrix. Since the initial estimate of the largest eigenvector now lies within the null space of the transformation the operation will not terminate in theory. In practice, this verified the operation of the code triggered by `ITERATE` and additionally demonstrated that the initial vector was not perturbed (remained within the null space of the transformation) after many iterations. The code was revised (not shown) by substituting the initial vector $[1, 0, \ldots, 0]^{\mathrm{T}}$ and no further difficulties were encountered.

# Appendix B

# Geometry of $N$-space and Zonotopes

## B.1 The Euclidean Geometry of $N$-space

Modified methods of basis construction give rise to new classes of reflectance solids. To assist in their exploration, a thumb-nail outline of terms is provided which define geometric features of Euclidean spaces of arbitrary dimension. For an in-depth treatment of regular polytopes and the geometric nature of $N$-dimensional Euclidean space, the reader is referred to (Coxeter, 1973) and (Grünbaum, 1967). A formal proof which demonstrates that the central sections of hypercubes form the zonotopes appeared concurrently in a popular text shortly after its colour-specific discovery here (Kaprass, 1992). For a more intuitive introduction to convex solids and (hyper)space, the reader is referred to (Critchlow, 1970), (Loeb, 1976), (Brisson, 1978), (Miyazaki, 1986).

Spaces of dimension higher than three are commonly called *hyperspace.* Here

an *n-space* means an $n$-dimensional Euclidean space. An $n$-dimensional solid lying in $n$-space is called an $n$-polytope. Common names for two- and three-dimensional polytopes are polygon and polyhedron, respectively. An $n$-polytope consists of (is bounded by) polytopes of lower increasingly reduced dimension $n-1$ to zero. The $(n-1)$ cells which bound the parent solid are called *facets.* The common names *face, edge* and *vertex* describe the 2-cell, 1-cell and 0-cell polytopes when they serve as bounding cells of a higher-dimensional solid. In the remainer of what follows merely convex polytopes are considered.

## B.2 Convex Space Partitions

Undifferentiated Euclidean space is convex, for the segment spanning any two points $A$ and $B$ lie within the space. Regions created by the successive partitioning of space preserve convexity, by induction: Introduction of a new partitioning plane preserves convexity, for if $A$ and $B$ lie to a common side of the cuting plane, so must all intermediate points along the segment $\overline{AB}$. A rigorous convexity proof (Coxeter, 1961) makes no guarantee that the polytope contains volume: the region of intersection may be empty. That is, the convex partitions may be semi-infinite, finite or empty.

The partitioning of convex $n$-space by (hyper)planes eventually gives rise to the $n$-dimensional polytopes. Further partitioning may create polytopes of reduced dimension $n-1$ through 0, ending with polygons, segments and vertices. No member of the series need be fully bounded: open sets such as rays and semi-open pyramids are also possible. Ultimately, the empty partition is created, which also satisfies complexity.

Any convex polytope can be constructed through the partitioning of convex

space. Here the cutting hyperplanes are chosen to contain the boundary facets of the solid. Algebraically, each $n-1$ facet lying in $n$ space has a normal vector defined by a plane equation. Algebraically, the hyperplane is a single (in)equality in a system of $n$ unknowns, producing a manifold having $n-1$ dimensions:

$$c_0 = (x_1, \cdots, x_n) \cdot (c_1, \cdots, c_n). \qquad [B.1]$$

The vector of coefficients $(c_1 \ldots c_n)$ define the orientation of the vector; the offset $c_0$ the distance to the origin when the normal vector is of unit length. To form a half-space partitioning the equality is replaced by an inequality, with sign indicating whether the origin lies within or without the convex halfspace.

## B.3 Multiple Partitions

Introducing additional partitions maintains geometric convexity. Algebraically, increasing the number of simultaneous equations to $k$ reduces the number of remaining degrees of freedom to $n-k$. The region of common intersection then takes on increasingly reduced dimension. As an example, the 3-space polyhedron's boundary consists of faces (2-cells), bound by single planes. Faces join along edges (1-cells) common to two defining planes. The vertices (0-cells) are the set of boundary points common to three defining planes.

For reflectance solids lying in an $m$ reflectance space of reduced dimension $m < k$ the system becomes overdetermined: there are more defining half-spaces than the dimension of the space. In this case the first $m$ planes enclose the solid, the remaining $k-m$ planes may or may not intersect. If intersecting, they will further sculpt its shape by intersection with facets already present. In the limit, a hyperplane just making contact with the solid will be coincident with a vertex (because of

convexity). Here $m - k$ becomes negative and the vertex is considered *degenerate.* (Lying on the surface of more than $m$ defining hyperplanes within the $m$ space. For instance, the vertices of the cube are not degenerate, as each is common to no more than three edges.) The presence of degenerate vertices often represents pathology to computational geometric algorithms seeking sub-quadratic running times (Preparata and Shamos, 1985).

Ultimately, an $n$-polytope is bounded by a set of vertices which are common to $n$ hyperplanes. The vertex set defines the polytope's *convex hull.* The polytope can be fully reconstructed from its convex hull, as the latter encodes the polytope's entire boundary information. In practice, a subset of vertices which is planar for which the remaining vertices lie to one side define the hyperplane for one facet of the solid. Enumeration of all such planes defines the solid in algebraic form, with features derived as needed by plane intersection. The first-principles procedural methods employed here have a running time $O(h^2)$ with $h$ the number of halfspaces. The virtue of the software (described in Appendix E) lies in the operational simplicity when operations are extended to spaces of higher dimension, wherein algebraic forms are extended easily, geometric data structures with great difficulty. The number of vertices defining the *convex hull* of the solid need not be closely related to the number of partitions, as may become excluded by halfspace (non)intersection.

Algorithms for polytope construction have been described (Chand and Kapur, 1970); related convex hull methods are known (Seidel, 1981), (Preparata, 1977) (Preparata, 1979), and have been surveyed (Matheiss and Rubin, 1980). Recent findings relate the problem to linear programming (Seidel, 1990). The newest algorithms (Chen, Hansen and Jaumard, 1991) which enumerate vertices in a convex hull reach asymptotic speeds provided few *degenerate* vertices exist. The unmodified reflectance solids described in Chapter 4 have at most one degenerate vertex:

Figure B.1: Regular Polytopes in $N$-space

the black apex. The modified solids introduce merely a conjugate white apex.

## B.4 Basic Polytopes

Any $n$-space contains at least three regular and convex solids (Coxeter, 1973). These are the *simplex,* the *cross-solid* and the *measure solid.* They generalize the concepts of tetrahedron, octahedron and cube (hexahedron) in three space. They are described in order. In §5.2.2 they are treated as a prototypical colour solids and the nature of their related basis explored. Representative solids appear in Figure B.1.

### B.4.1 Simplex

The (regular) simplex may be formed from the convex hull of $n+1$ (equidistant) vertices; by duality (Holden, 1971) it has $n+1$ opposing facets. It is the minimal solid to enclose volume in any $n$-space. This follows because at least $n$ vertices are required to define one boundary cell (e.g., three vertices in three-space define one triangular face, a polygonal facet having fewest vertices). When $n+1$ independent

points define a convex null then $n + 1$ distinct subsets of $n$ vertices locate the bounding facets which oppose the unchosen vertex. The simplex is self-dual and extends the sequence which begins with (line) segment, triangle, tetrahedron.

### B.4.2 Cross Solid

The cross-solid is the convex hull formed by a vertex pair $(\pm 1, 0, \ldots, 0)$ permuted cyclically. Each pair defines a segment along a Cartesian axis through the origin, forming the cross. The solid has $2n$ vertices and $2^n$ cells. In 3-space, the cross solid is the octahedron. When presented on-axis, its dual relation to the cube and internal cross are evident in Figure B.1.

### B.4.3 Measure Solid

The measure solid generalizes the rectangle and parallelepiped. It is the *hypercube* when regular, forming the dual of the cross solid. Generally, it has $2^n$ vertices at $(\pm c_1, \pm c_2, ..., \pm c_n)$ with the "$\pm$" signs taken independently. Its facets are $(n-1)$-dimensional measure solids arranged pairwise about the Cartesian axis. That is, each coordinate axis is normal to two facets, and defines each of the orthogonal zones. (That is, the set of paired planes is orthogonal.) The solid can tile (pack without void) $n$-space. The packing of space with unit $n$-hypercubes provides a measure of unit volume for the space, hence its name.

## B.5 Parallelotopes

The hypercube's opposing cells may be sheared (moved laterally within their defining hyperplanes) to produce a *parallelotope,* a generalization of the parallelepiped.

This action leaves unchanged the $n$-volume (Eves, 1987), plus the number of cells of dimension $n-1$ through zero. The parallelotope may be regarded as a distortion of the space's basis, whose vectors are independent but no longer orthogonal; the geometric features appear in distorted yet complete interrelation. This suggests a means for recording the shape of a parallelotope which is more compact than recording its $2^n$ vertices. Instead, the vertices of its dual are recorded as a geometric figure of $n$ segments which span the $(2n)$ opposing vertices. This forms a *cross,* explaining the use of the term and the related "cross-solid" of the dual figure.

All solids discussed, save the simplex, are composed of an even number of opposing, parallel facets, verified by showing that the normal vector of paired facets is identical (to within a change of sign).

## B.6 Zonotopes

Parallel, bracketing hyperplanes form an interval of inclusion. The interval is bound along one axis yet otherwise of infinite extent. It is called a *zone.* The volume of intersection common to a set of zones is a condition necessary in the definition of *zonotopes,* seen in Figure B.2(a). The definition is sufficient should all zones contain a common point equidistant from the bracketing parallel planes, as in Figure B.2(b). When this point lies at the polytope's centroid, it establishes a *centre of inversion.* A cross may be constructed as above which records the thickness and orientation of each zone, as in Figure B.2(c). When all arms of the cross are bisected by the common midpoint, the cross formed is called a *eutectic star* by Coxeter[1]. This figure is uniquely determines the zonotope.

[1] *op cit.*

Figure B.2: Creation of Zonotopes

Zonotopes may have up to $f_z = z(z-1)$ faces, where $z$ is the number of defining zones, as any zone may interpenetrate any other save for itself. The number of edges and vertices are of similar order: $e_z = 2z(z-1)$ and $v_z = z(z-1)+2$ for zonohedra with parallelogram faces (Ball and Coxeter, 1974). Note that $f_e - 2 = f_z + v_z$ by Euler's formula.

To summarize,

$$\text{zonotopes} \subseteq \text{parallelotopes} \subseteq \text{convex polytopes}. \qquad [B.2]$$

Because the simplex is not a parallelotope, it is not a zonotope. The measure solid is the prototypical zonotope, having a eutectic star composed of orthogonal segments having equal lengths all coincident at their midpoints. This relation provides the bridge which helps link the exploration of geometric reflectance solids and linear algebraic functions. As an example, the solids depicted in §4.2.9 in 3-space exhibit "belts" and "creases". Since these are one-dimensional structures, they are common to $3-1$ defining planes. Since the planes define the limits of reflectance, all points along these features exhibit reflectance functions which reach their maximum and/or minimum values on the interval $[0 \ldots 1]$ at exactly two points.

# Appendix C

# Spectroradiometry Procedures

In general, spectroradiometric data is presented as column vectors with an optional "0th" prefacing wavelength vector. An extensive matrix software library was as traditional tools such as MATLAB were unable to accommodate the large matrices created in connection with SVD. The custom library was created instead. It provided three useful features. First, it allows for run-time sizing of the array, while preserving (much as is possible) annotations on the data set, its rows and columns, both across file read and matrix operations. In addition, it allows the creation of fictitious rows formed by interpolation of adjacent rows having bracketing wavelength value (in the 0th column). By this operations such as the inner-project of matrix multiplication may be extended to better resemble the integration of the product of two functions indexed by wavelength. The library and its construction are the subject of Appendix E.

A set of colorimetric tools were constructed atop the library, appearing in Table E.2. The scan line interpolation function allows for the straightforward implementation of such operations as reflectance function onto chromaticity coordinate

mapping (tool SPECCOORD). These in turn may read in annotated CIE-defined tables of values be virtue of the underlying matrix library. These tables appear verbatim in Appendix G.

## C.1 Reflectance to Device Coordinate Mapping

The non-linear mapping between (three) reflectance coordinates and (three) print values employs conventional colour correction software. Traditionally, such software maps between additive colour spaces (RGB, XYZ or LUV) whose components are non-negative. Reflectance coordinates do not have this restriction. Conventional mapping software may be easily generalized to account for this condition; its proper treatment was set forth as a design requirement by the author in a system created at Waterloo through the student co-op program in anticipation of this reuse with reflectance-based data. A custom tool was eventually created as the general "gamut mapping" laboratory software evolved in other directions (and because of the high turnover of student programmers, which complicated maintenance and documentation). Based on tri-linear interpolation (Appendix H) a particularly compact tool, INVERTRGB was created. Its C-language source appears below.

```
/*
 * invertrgb.c -- map beta back in to R G B coordinates.
 *
 * input: data records of the form "b1, b2, b3, ...extra-in..."
 * and an inversion list of the form "r g b b1 b2 b3 ...extratab..."
 * find the linear interp of the rgb input which matches the data beta and
 * output "r g b ...extra-in..."
 */

#include <stdio.h>
#include <math.h>
#include <matlib.h>
```

```
#define PROGNAME "invertrgb"

#define USAGE "\nusage: invertrgb samplefile invertbasis\n\
reexpress beta (cols 1..3) of samples as (cols 1..3) of basis, such that\n\
the linear interp of the basis onto (cols 4..6) matches the input\n\
 -f   merely do the forward linear interpolation\n\
 -p N evaluate only for records with fourth col == N, else transparent\n"

main(argc, argv)
  char **argv;
  {
  int fflag, pflag, r, c, i;
  char *samname, *basname;
  mat sam, bas, rhs, lhs;
  mat p0, p1, p2, p3, p4, p5, p6, p7;
  mat p01, p23, p45, p67;
  mat p03, p47;
  mat p07;
  double s01, s23, s45, s67;
  double s03, s47;
  double s07;
/*
 * defaults
 */
  fflag = 0;
  pflag = -1;
/*
 * parse the command line
 */
  for(i=1; i<argc; i++)
    {
    if (argv[i][0] == '-')
      {
      if (argv[i][1] == '\0') err("lone hyphen found");
      if (argv[i][2] != '\0') err("switch has more than one character");
      switch (argv[i][1])
        {
        case 'F':
        case 'f': fflag = 1; break;
        case 'P':
        case 'p': i++;
             if (i == argc) err("trailing -p is illegal");
             pflag = atoi(argv[i]);
             break;
        case 'U':
        case 'u': err(USAGE); break;
```

```
        default: err("unknown switch"); break;
        }
      }
    else
      {
      if (samname)
        {
        if (basname) err("too many cmd line files");
        basname = argv[i];
        }
      else samname = argv[i];
      }
    }
/*
 * open, check
 */
  sam = matread(samname);
  bas = matread(basname);
  if (C(sam) < 3) err("sample set requires at least three columns");
  if (C(bas) < 6) err("basis set requires at least six columns");
/*
 * -f switch "hack" - just reverse entire domain and range!
 */
  if (fflag)
    {
    for (r = 0; r<R(bas); r++)
      {
      double t0, t1, t2;
      t0 = EL(bas,r,0);
      t1 = EL(bas,r,1);
      t2 = EL(bas,r,2);
      EL(bas,r,0) = EL(bas,r,3);
      EL(bas,r,1) = EL(bas,r,4);
      EL(bas,r,2) = EL(bas,r,5);
      EL(bas,r,3) = t0;
      EL(bas,r,4) = t1;
      EL(bas,r,5) = t2;
      }
    }
/*
 * now process
 */
  p0 = matalloc(1, 8);
  p1 = matalloc(1, 8);
  p2 = matalloc(1, 8);
  p3 = matalloc(1, 8);
  p4 = matalloc(1, 8);
```

```
  p5 = matalloc(1, 8);
  p6 = matalloc(1, 8);
  p7 = matalloc(1, 8);
  p01 = matalloc(1, 8);
  p23 = matalloc(1, 8);
  p45 = matalloc(1, 8);
  p67 = matalloc(1, 8);
  p03 = matalloc(1, 8);
  p47 = matalloc(1, 8);
  p07 = matalloc(1, 8);
  rhs = matalloc(1, 8);
  lhs = matalloc(1, 8);
/* for each record */
  for (r = 0; r<R(sam); r++)
    {
    int t;
    t = 1;
    if ((pflag < 0) || (pflag == EL(sam,r,3)))
      {
/* find the corners of the bracketing cube */
      t &= findnear(p0, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 0);
      t &= findnear(p1, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 1);
      t &= findnear(p2, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 2);
      t &= findnear(p3, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 3);
      t &= findnear(p4, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 4);
      t &= findnear(p5, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 5);
      t &= findnear(p6, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 6);
      t &= findnear(p7, bas, EL(sam,r,0), EL(sam,r,1), EL(sam,r,2), 7);
  /* blend down */

      s01 = (EL(sam,r,2) - EL(p0,0,5)) / (EL(p1,0,5) - EL(p0,0,5));
      s23 = (EL(sam,r,2) - EL(p2,0,5)) / (EL(p3,0,5) - EL(p2,0,5));
      s45 = (EL(sam,r,2) - EL(p4,0,5)) / (EL(p5,0,5) - EL(p4,0,5));
      s67 = (EL(sam,r,2) - EL(p6,0,5)) / (EL(p7,0,5) - EL(p6,0,5));

      matscale(rhs,p0,1.0-s01);matscale(lhs,p1,s01);matadd(p01, rhs,lhs);
      matscale(rhs,p2,1.0-s23);matscale(lhs,p3,s23);matadd(p23, rhs,lhs);
      matscale(rhs,p4,1.0-s45);matscale(lhs,p5,s45);matadd(p45, rhs,lhs);
      matscale(rhs,p6,1.0-s67);matscale(lhs,p7,s67);matadd(p67, rhs,lhs);

      s03 = (EL(sam,r,1) - EL(p01,0,4)) / (EL(p23,0,4) - EL(p01,0,4));
      s47 = (EL(sam,r,1) - EL(p45,0,4)) / (EL(p67,0,4) - EL(p45,0,4));

      matscale(rhs,p01,1.0-s03);matscale(lhs,p23,s03);matadd(p03,rhs,lhs);
      matscale(rhs,p45,1.0-s47);matscale(lhs,p67,s47);matadd(p47,rhs,lhs);

      s07 = (EL(sam,r,0) - EL(p03,0,3)) / (EL(p47,0,3) - EL(p03,0,3));
```

```
      matscale(rhs,p03,1.0-s07);matscale(lhs,p47,s07);matadd(p07,rhs,lhs);
/*
 * overwrite first three values in sample point
 */
      for (c=0; c<3; c++) EL(sam,r,c) = EL(p07,0,c);
      }
/*
 * print current record (transparent or inverted) with other trailing data
 */
    for (c=0; c<C(sam); c++) printf(" %G", EL(sam,r,c));
    printf(t ? "\n" : " #out of gamut\n");
    }
  }

#define MAXSTART 1e10

findnear(mrow, m, s1, s2, s3, code)
  mat mrow, m;
  double s1, s2, s3;
  int code;
  {
  double b1, b2, b3, dist, maxdist;
  int r, c, rec, code2;
  maxdist = MAXSTART;
  for (r=0; r<R(m); r++)
    {
    code2 = 0;
    b1 = EL(m,r,3);
    b2 = EL(m,r,4);
    b3 = EL(m,r,5);
    if (b1 > s1) code2 += 4;
    if (b2 > s3) code2 += 2;
    if (b3 > s3) code2 += 1;
    if (code == code2)
      {
      dist = (s1-b1)*(s1-b1)+(s2-b2)*(s2-b2)+(s3-b3)*(s3-b3);
      if (dist < maxdist)
        {
        maxdist = dist;
        rec = r;
        }
      }
    }
  if (maxdist == MAXSTART) return(0);
  for (c=0; c<C(mrow); c++) EL(mrow,0,c) = EL(m,rec,c);
  return(1);
```

```
  }

err(st)
  char *st;
  {
  fprintf(stderr, "\n%s: %s\n", PROGNAME, st);
  exit(1);
  }
```

# Appendix D

# Installation of the Kodak 6500 Printer

Cable installation required a VAX8600 to Centronix parallel interface. Information was solicited from the net, a DECUS VAX SIG newsletter was provided by Tom Allebrandi of Advanced Computer Consulting, Inc. 804/9977-4272. The memorandum is presented below.

```
                        INPUT/OUTPUT 349

Caption:  Centronics interface to DMF-32 -- reply to I/O # 321

Message:  We have successfully interfaced the DMF-32 to our  own
          printers,  which  are  Centronics inteface compatible.
          You may only need to  make  a  cable  and  strap  your
          printer   accordingly.    If   your  printer  has  the
          "standard" 36-pin Centronics connector (same as  Epson
          MX-80  and  others),  you  can  make  a cable as shown
          below;  otherwise you will have to locate the  signals
          on your printer connector and make a custom cable.

          DMF32                                       Centronics
```

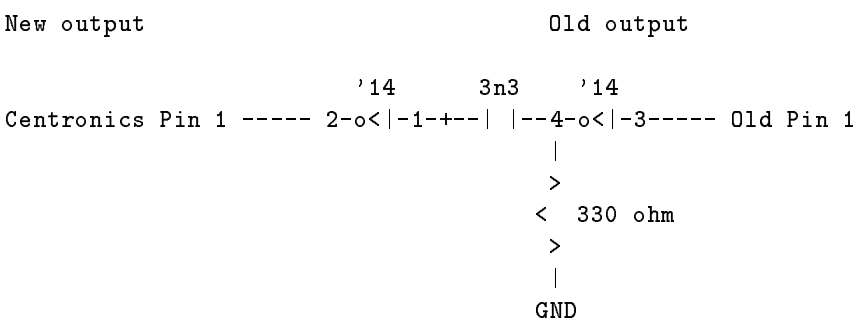
```
 J14               Signal Name            Connector
 --                -----                      -----
  8                Strobe                       1
 26                Data 1                       2
 20                Data 2                       3
 22                Data 3                       4
  1                Data 4                       5
 24                Data 5                       6
 23                Data 6                       7
  5                Data 7                       8
  6                Data 8                       9
 18                Busy (demand)               11
 12                Select (online)             13
 17                Fault                       32
 14                POn (conn VFY)              16
 30                0V and Shield               16

        Appendix C in the DMF32 User's Guide (EK-DMF32-UG-002)
        is  incorrect;  you  should  mark  J14-37  as PRINTER
        SPARE,  and  J14-8  as  PRINTER  STROBE.  Note  that
        ACKnowledge  and  PE (paper empty) are not used.  Your
        printer should be strapped to go off-line  (deselected,
        local) on a fault condition;  otherwise VMS won't tell
        you anything when paper runs out.

        Additional printer strapping information:

        Busy level = low (0)
        Strobe level = high (1)
        Strobe trigger = trailing (falling) edge

        For a long cable, if  the  printer  allows,  the  data
        lines  should be sampled 300-500 nanoseconds after the
        trailing edge of the strobe pulse.

Contact: Carl Houseman
         GENICOM Corporation
         1 General Electric Drive
         Waynesboro, VA 22980
         (703) 949-1323

Date:    November 18, 1984
```

From this a final cable was created by Mike Gore, ICR at the University of Waterloo. Because of a signal polarity reversal as provided by the VAX a gate was

required to provide signal inversion for one pin. Because power is not available on the cable and an outboard transformer prohibitive, a low-power CMOS inverter was installed within the cable hood which draws its power from a +5 (logic high) signal elsewhere on the cable, seen below.

```
================================================================
              Kodak SV6500 "Centronics" interface
              Mike Gore         November 12, 1990
================================================================

Centronics                        SV6500
Pin     Signal                    Signal        Pin

 1*     /Strobe    ----------------> /Ext Write  1*
 2      Data 0     ----------------->Data 0      2
 3      Data 1     ----------------->Data 1      3
 4      Data 2     ----------------->Data 2      4
 5      Data 3     ----------------->Data 3      5
 6      Data 4     ----------------->Data 4      6
 7      Data 5     ----------------->Data 5      7
 8      Data 6     ----------------->Data 6      8
 9      Data 7     ----------------->Data 7      9
10      /ACK       <---------------- /FS ACK    10
11      Busy       <---------------- Busy       11
12*     Page End   <----------------             12*
13*     Select     <----------------             13*
14      /Auto feed ----------------> /Ext Write 14
31*     /Init      <----------------             31*
32*     /Error     <---------------- /Error     32*
36      /Select In ----------------->            36
16,19-30,33 GND    ----------------- GND        16,19-30,33


Kodak printer special considerations

Notes:
    /Strobe MUST be between .5 and 2.5 micro seconds in length.

    Pin 12,13,31,32 have no names in the SV6500 docs but are
    assumed to be used.

Serial/Centronics converter mods:
```

```
   1) Pin 16,33 &36(/Select) have been grounded in the converter.
   2) Pin 31(/Init) has been tied to VCC with a 1.2K resistor
   3) Pin 1 has been cut and the following circuit added:
      Notes:
         This circuit reduces the write strobe to about 1.4 uS.
         '14 = 74ls14 inverter
         3n3 = 3.3 * 10exp-9 Farad
```



# Appendix E

# C-language Libraries and Tools

The library of C-language software tools is described in Table E.1. More detailed descriptions are available for most tools by invoking them with the -u (usage) option.

A C library supporting matrices of arbitrary length was created. The routines are listed Table E.2. Names are essentially self-explanatory; those ending with "a" indicate allocation. For instance, matrices `a` and `b` may be added to produce `c` using `matadd(c,a,b)`; but `matadda(a,b)` creates the sum and returns the result in a new matrix. The tools in all cases check for conformality and other erroneous conditions.

Examples of use of this matrix library appear in the source code presented in Appendices A and C.

| Program | Description |
|---|---|
| BLACKTORGB | find rgb triples for colour temp |
| CIECONV | stand-alone C routines to map spectra |
| CIEGRID | draw chromaticity diagram on Ikonas |
| CIEINTERP | interpolate (wave) lines into (wave,X,Y,Z) lines |
| CIEUV | give a hue name to an x',y' or X,Y,Z value |
| FITBASIS | given an input and basis file, reexpress former on latter |
| FIXUP | make norm again |
| IMSPECTRUM | build a scan line of some spectral intensity |
| INVERTRGB | map beta back in to R G B coordinates |
| LEMPLOT | scatter plots or connected line plots of data sets |
| LEMTEXT | map text files into lemming editor text documents |
| LOCALSVD | map coordinates into basis elements |
| MATMERGE | invoke a multi-file read (matrix augmentation) then print |
| MATMUL | given an input and basis file, reexpress former on latter |
| META4D | find 3D space of [0..1] curve metamers |
| ORTBASIS | given an input file(s) generate the first N eigenvectors |
| SOLVE3X3 | solve an general 3x3 linear system |
| SPECBASIS | like "fitbasis" but integrates radiometer code |
| SPECCOORD | compute x,y,z,x',y',l,u,v,D for (n) spectra, plus colour |
| SPECFIX | subtract black, normalize by white |
| SPECINTERP | remap input data to conform to output basis |
| SPECNEAR | compute x,y,z,x',y',l,u,v,D for (n) spectra, plus colour |
| SPECNEARTAB | tabulate list of nearest bases |
| SUBSPEC | image of all rectangular filters across the visible spectrum |
| SV6500 | Centronix printing with tty output and parallel adaptor |
| SVDLOCUS | plot 8x8x8 colour svd interpolant |
| TRIGBASIS | form (near) orthgonal basis tables for viewing |
| WHITENORM | divides row elements [1..$n$] of the input by the aux input |

Table E.1: Software Tools

| | | | | |
|---|---|---|---|---|
| matadd | matadda | matalloc | mataudit | mataugment |
| matbounds | matcalloc | matcheck | matcieinp | matcolmax |
| matcolmin | matcols | matcolsa | matcolvec | matcolveca |
| matconf | matconst | matconsta | matcopy | matcopya |
| mateigen | mateigenbig | matequal | matfree | matglobals |
| matgram | matgrama | matinvert | matmul | matmula |
| matname | matread | matreg | matreturn | matrowinp |
| matsuba | matsyminv | mattrace | mattrans | mattransa |
| matvalinp | matvecdot | matveclen | matvecmax | matvecunit |
| matwrite | | | | |

Table E.2: MATLIB Matrix Routines

# Appendix F

# RGB Transformation Matrices

## F.1 XYZ-RGB Conversions

These conversions map between tristimulus vectors and pixel illumination coordinates for the specified monitor chromaticities. In general, these are specified as (x',y') pairs for each of the RGB phosphors in a triad. The value $z'$ is easily deduced since $x' + y' + z' = 1$. Arranged as a matrix, these form a transformation matrix which specifies the tristimulus values recorded under excitation (generation) of each phosphor. This is best described given an example.

### F.1.1 Iris 3020 series workstations

The RGB transformation matrices used by the Iris mapping software are based on the Hitachi colour monitor model #CM2073ASG-517 used on the Silicon Graphics Iris Workstation (3020 series) used in support of the research. Chromaticities were provided by Mr. D. Christopher Dunlap of SGI Hardware Product Support in October 1988 and appear in Table F.1.

| | Chromaticity | | Phosphor | |
|---|---|---|---|---|
| Colour | x | y | Characteristic | Persistence |
| Red | 0.610 | 0.342 | medium | 1.2 msec |
| Green | 0.298 | 0.588 | short | 300 usec |
| Blue | 0.151 | 0.064 | short | 250 usec |

Table F.1: Iris 4D Phosphor Chromaticities and Persistence Characteristics

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 0.6100 & 0.2980 & 0.1510 \\ 0.3420 & 0.5880 & 0.0640 \\ 0.0480 & 0.1140 & 0.7850 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} . \qquad [F.1]$$

thus defining the XYZ to RGB transformation.

Note that the transposition of the tabular data means that (for example) post-multiplication by the "red" row vector [100] $\mathbf{T}$ recovers the tabulated chromaticities $[x'y'z']^{\mathrm{T}}$ for that phosphor, also as a row. Note that the conditions for unit row sums is easily verified by the identity [111] $\mathbf{T}$ = [111]. That is, the ones vector is a left-eigenvector[1]. This algebraic form is strongly similar to the condition for unit row sums in the MSSM method.

Matrix inversion finds the chromaticity to pixel coordinate conversion. This is the more useful of the two: their employ provides a means of device-independent colour specification using tristimulus (illumination) coordinates:

[1]Traditionally, the right-hand form $\mathbf{Mv} = \mu\mathbf{v}$ is used. This left-hand solution, analogous to the (right-hand) solution of $\mathbf{M}'$ preserves eigenvalues, including the unit, but destroys the accompanying eigenvector having constant elements.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.2754 & -1.0855 & -0.3492 \\ -1.3293 & 2.3621 & 0.0631 \\ 0.0539 & -0.2767 & 1.2861 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad [F.2]$$

Analogous methods may be used with other published phosphor chromaticities.

### F.1.2 NTSC

The NTSC reccommended chromaticities are relatively spectrally pure values which are seldom approximated in practice. Instead, "rare-earth" reds having reduced purity are often substituted as they have increased (perceptual) brightness.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} .67 & .21 & .14 \\ .33 & .71 & .08 \\ 0.0 & .08 & .78 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad [F.3]$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.73 & -.48 & -.26 \\ -.81 & 1.65 & -.02 \\ .08 & -.17 & 1.28 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad [F.4]$$

### F.1.3 Spectral RGB

A larger (and less probable) gamut is based on the "maximal" gamut having chromaticities on the spectrum locus for RGB sources at 450nm, 500nm and 600nm.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} .63 & .30 & .16 \\ .37 & .69 & .02 \\ 0.0 & .01 & .82 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad [F.5]$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.13 & -.92 & -.40 \\ -1.14 & 1.94 & .17 \\ .01 & -.02 & 1.21 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad [F.6]$$

### F.1.4 Electrohome

A lab commonly used in Waterloo's Computer Graphics Laboratory, manufactured locally by a national corporation.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} .62 & .21 & .15 \\ .33 & .675 & .06 \\ .05 & .115 & .79 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad [F.7]$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.93 & -.545 & -.325 \\ -.945 & 1.77 & .045 \\ .015 & -.222 & 1.28 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad [F.8]$$

### F.1.5 Aydin

The Aydin is a research-grade laboratory monitor.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} .65 & .3 & .15 \\ .3 & .6 & .1 \\ .05 & .1 & .75 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad [F.9]$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 2.00 & -.955 & -.273 \\ -1.00 & 2.18 & -.091 \\ 0.0 & -.227 & 1.36 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad [F.10]$$

## F.2 YIQ-RGB conversion

Finally, the YIQ encoding matrices are presented. Orthogonality of the columns gives rise to an inverse having units in the first column:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & .587 & 0.114 \\ 0.599 & -.2773 & -.3217 \\ .213 & -.5251 & .3121 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \qquad [F.11]$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0.9514 & 0.6154 \\ 1.0 & -.2702 & -.6438 \\ 1.0 & -1.104 & 1.7009 \end{bmatrix} \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \qquad [F.12]$$

# Appendix G

# CIE and Standard Tables

Below are the spectral power distribution files used in connection with this research. The data files were transcribed from well-known texts; any errors are the responsibility of the author.

## G.1 Spectral Power Distributions

```
# Spectral Power Distributions for CIE
# sources (Abridged); wavelengths in nm.

#lambda D50     D65     A
#       -----   -----   -----
#
380     24.49   50.0    9.80
385     27.18
390     29.87   54.6    12.09
395     39.59
400     49.31   82.8    14.71
405     52.91
410     56.51   91.5    17.68
415     58.27
420     60.03   93.4    20.99
```

| | | | |
|---|---|---|---|
| 425 | 58.93 | | |
| 430 | 57.82 | 86.7 | 24.67 |
| 435 | 66.32 | | |
| 440 | 74.82 | 104.9 | 28.70 |
| 445 | 81.04 | | |
| 450 | 87.25 | 117.0 | 33.09 |
| 455 | 88.93 | | |
| 460 | 90.61 | 117.8 | 37.81 |
| 465 | 90.99 | | |
| 470 | 91.37 | 114.9 | 42.87 |
| 475 | 93.24 | | |
| 480 | 95.11 | 115.9 | 48.24 |
| 485 | 93.54 | | |
| 490 | 91.96 | 108.8 | 53.19 |
| 495 | 93.84 | | |
| 500 | 95.72 | 109.4 | 59.86 |
| 505 | 96.17 | | |
| 510 | 96.61 | 107.8 | 66.06 |
| 515 | 96.87 | | |
| 520 | 97.13 | 104.8 | 72.50 |
| 525 | 99.61 | | |
| 530 | 102.10 | 107.7 | 79.13 |
| 535 | 101.43 | | |
| 540 | 100.75 | 104.4 | 85.95 |
| 545 | 101.54 | | |
| 550 | 102.32 | 104.0 | 92.91 |
| 555 | 101.16 | | |
| 560 | 100.00 | 100.0 | 100.0 |
| 565 | 98.87 | | |
| 570 | 97.74 | 96.3 | 107.18 |
| 575 | 98.33 | | |
| 580 | 98.92 | 95.8 | 114.44 |
| 585 | 96.21 | | |
| 590 | 93.50 | 88.7 | 121.73 |
| 595 | 95.59 | | |
| 600 | 97.69 | 90.0 | 129.04 |
| 605 | 98.48 | | |
| 610 | 99.27 | 89.6 | 136.35 |
| 615 | 99.16 | | |
| 620 | 99.04 | 87.7 | 143.62 |
| 625 | 97.38 | | |
| 630 | 95.72 | 83.3 | 150.84 |
| 635 | 97.29 | | |
| 640 | 98.86 | 83.7 | 157.98 |
| 645 | 97.26 | | |
| 650 | 95.67 | 80.0 | 165.03 |
| 655 | 96.93 | | |

```
660     98.18   80.2    171.96
665     100.60
670     103.00  82.3    178.77
675     101.07
680     99.13   78.3    185.43
685     93.26
690     87.38   69.7    191.93
695     89.49
700     91.60   71.6    198.26
705     92.25
710     92.89   74.3    204.41
715     84.87
720     76.85   61.6    210.36
725     81.68
730     86.51   69.9    216.12
735     89.55
740     92.58   75.1    221.67
745     85.40
750     78.23   63.6    227.00
755     67.96
760     57.69   46.4    232.12
765     70.31
770     82.92   66.8    237.01
775     80.60
780     78.27   63.4    241.68
```

## G.2 Tristimulus Vectors (Tables)

The following are the tristimulus vectors for the CIE 1931 Standard Observer (2° field of view). The same disclaimers apply. In Table G.1 (following), the chromaticities of 35mm Ektachrome slide film are presented. These were the entries used in the creation of the colour plate (Foley et al., 1990) described in §3.2.2.

```
# L(nm) X       Y        Z
#
380     0.0014  0.0000  0.0065
385     0.0022  0.0001  0.0105
390     0.0042  0.0001  0.0201
395     0.0076  0.0002  0.0362
```

| | | | |
|---|---|---|---|
| 400 | 0.0143 | 0.0004 | 0.0679 |
| 405 | 0.0232 | 0.0006 | 0.1102 |
| 410 | 0.0435 | 0.0012 | 0.2074 |
| 415 | 0.0776 | 0.0022 | 0.3713 |
| 420 | 0.1344 | 0.0040 | 0.6456 |
| 425 | 0.2148 | 0.0073 | 1.0391 |
| 430 | 0.2839 | 0.0116 | 1.3856 |
| 435 | 0.3285 | 0.0168 | 1.6230 |
| 440 | 0.3483 | 0.0230 | 1.7471 |
| 445 | 0.3481 | 0.0298 | 1.7826 |
| 450 | 0.3362 | 0.0380 | 1.7721 |
| 455 | 0.3187 | 0.0480 | 1.7441 |
| 460 | 0.2908 | 0.0600 | 1.6692 |
| 465 | 0.2511 | 0.0739 | 1.5281 |
| 470 | 0.1954 | 0.0910 | 1.2876 |
| 475 | 0.1421 | 0.1126 | 1.0419 |
| 480 | 0.0956 | 0.1390 | 0.8130 |
| 485 | 0.0580 | 0.1693 | 0.6162 |
| 490 | 0.0320 | 0.2080 | 0.4652 |
| 495 | 0.0147 | 0.2586 | 0.3533 |
| 500 | 0.0049 | 0.3230 | 0.2720 |
| 505 | 0.0024 | 0.4073 | 0.2123 |
| 510 | 0.0093 | 0.5030 | 0.1582 |
| 515 | 0.0291 | 0.6082 | 0.1117 |
| 520 | 0.0633 | 0.7100 | 0.0782 |
| 525 | 0.1096 | 0.7932 | 0.0573 |
| 530 | 0.1655 | 0.8620 | 0.0422 |
| 535 | 0.2257 | 0.9149 | 0.0298 |
| 540 | 0.2904 | 0.9540 | 0.0203 |
| 545 | 0.3597 | 0.9803 | 0.0134 |
| 550 | 0.4334 | 0.9950 | 0.0087 |
| 555 | 0.5121 | 1.0000 | 0.0057 |
| 560 | 0.5945 | 0.9950 | 0.0039 |
| 565 | 0.6784 | 0.9786 | 0.0027 |
| 570 | 0.7621 | 0.9520 | 0.0021 |
| 575 | 0.8425 | 0.9154 | 0.0018 |
| 580 | 0.9163 | 0.8700 | 0.0017 |
| 585 | 0.9786 | 0.8163 | 0.0014 |
| 590 | 1.0263 | 0.7570 | 0.0011 |
| 595 | 1.0567 | 0.6949 | 0.0010 |
| 600 | 1.0622 | 0.6310 | 0.0008 |
| 605 | 1.0456 | 0.5668 | 0.0006 |
| 610 | 1.0026 | 0.5030 | 0.0003 |
| 615 | 0.9384 | 0.4412 | 0.0002 |
| 620 | 0.8544 | 0.3810 | 0.0002 |
| 625 | 0.7514 | 0.3210 | 0.0001 |
| 630 | 0.6424 | 0.2650 | 0.0000 |

| | | | |
|---|---|---|---|
| 635 | 0.5419 | 0.2170 | 0.0000 |
| 640 | 0.4479 | 0.1750 | 0.0000 |
| 645 | 0.3608 | 0.1382 | 0.0000 |
| 650 | 0.2835 | 0.1070 | 0.0000 |
| 655 | 0.2187 | 0.0816 | 0.0000 |
| 660 | 0.1649 | 0.0610 | 0.0000 |
| 665 | 0.1212 | 0.0446 | 0.0000 |
| 670 | 0.0874 | 0.0320 | 0.0000 |
| 675 | 0.0636 | 0.0232 | 0.0000 |
| 680 | 0.0468 | 0.0170 | 0.0000 |
| 685 | 0.0329 | 0.0119 | 0.0000 |
| 690 | 0.0227 | 0.0082 | 0.0000 |
| 695 | 0.0158 | 0.0057 | 0.0000 |
| 700 | 0.0114 | 0.0041 | 0.0000 |
| 705 | 0.0081 | 0.0029 | 0.0000 |
| 710 | 0.0058 | 0.0021 | 0.0000 |
| 715 | 0.0041 | 0.0015 | 0.0000 |
| 720 | 0.0029 | 0.0010 | 0.0000 |
| 725 | 0.0020 | 0.0007 | 0.0000 |
| 730 | 0.0014 | 0.0005 | 0.0000 |
| 735 | 0.0010 | 0.0004 | 0.0000 |
| 740 | 0.0007 | 0.0002 | 0.0000 |
| 745 | 0.0005 | 0.0002 | 0.0000 |
| 750 | 0.0003 | 0.0001 | 0.0000 |
| 755 | 0.0002 | 0.0001 | 0.0000 |
| 760 | 0.0002 | 0.0001 | 0.0000 |
| 765 | 0.0001 | 0.0000 | 0.0000 |
| 770 | 0.0001 | 0.0000 | 0.0000 |
| 775 | 0.0001 | 0.0000 | 0.0000 |
| 780 | 0.0000 | 0.0000 | 0.0000 |

| Exp # | X' | Y' | Hue | f/stop |
|---|---|---|---|---|
| 18 | 0.333 | 0.333 | dmax | 0 |
| 42 | 0.313 | 0.313 | white | 2.8 |
| 44 | 0.298 | 0.291 | white | 5.6 |
| 02 | 0.272 | 0.277 | white | 11 |
| 04 | 0.225 | 0.233 | white | 22 |
| 11 | 0.350 | 0.362 | dmin | (inf) |
| 21 | 0.688 | 0.309 | red | 2.8 |
| 23 | 0.689 | 0.308 | red | 5.6 |
| 25 | 0.668 | 0.321 | red | 11 |
| 27 | 0.594 | 0.348 | red | 22 |
| 28 | 0.322 | 0.631 | green | 2.8 |
| 30 | 0.296 | 0.659 | green | 5.6 |
| 32 | 0.290 | 0.649 | green | 11 |
| 34 | 0.303 | 0.581 | green | 22 |
| 35 | 0.149 | 0.045 | blue | 2.8 |
| 37 | 0.150 | 0.041 | blue | 5.6 |
| 39 | 0.150 | 0.047 | blue | 11 |
| 41 | 0.154 | 0.071 | blue | 22 |
| 05 | 0.350 | 0.145 | magenta | 2.8 |
| 08 | 0.173 | 0.243 | cyan | 2.8 |
| 12 | 0.507 | 0.481 | yellow | 2.8 |

Table G.1: Normalized Ektachrome Film Chromaticity (source E)

# Appendix H

# Linear Interpolation

The linear interpolation step estimates the fractional change of output function (ordinate) between two samples as directly proportional to the fractional distance between the bracketing input samples. In the one-dimensional case this is

$$\begin{aligned} I_m &= (1-\alpha)I_a + \alpha I_b \\ P(I_m) &= P((1-\alpha)I_a) + P(\alpha I_b) \\ P(I_m) &= (1-\alpha)P(I_a) + \alpha P(I_b) \\ P((1-\alpha)I_a + \alpha I_b) &= (1-\alpha)P(I_a) + \alpha P(I_b), \end{aligned} \qquad [H.1]$$

in which $\alpha$ indicates the fractional distance along the $j$th axis. The operation is easily generalized to higher dimension by composition. The resulting surface is unique sheet hyperbola independent of the order of axis composition. Thus, the continuous approximation is continuous, though slope discontinuities (cusps) are possible at the sample points themselves. Derivations similar to the above with accompanying C-language code have recently appeared in a widely accessible text (Hill, 1994).

Linear interpolation fully characterizes an idealized printer having linear behaviour by recording the print densities occurring for the eight Cartesian primaries.

Real printers are not linear, but a continuous change in output as a function of input implies the existence of a multivariate Taylor expansion. Given a sufficiently small neighborhood (and a large gamut table) the printer's behaviour is linear as the quadratic and higher order terms quickly tend to zero. Under these conditions of local use the method provides a highly accurate estimate.

Methods of printer modeling using linear interpolation of been explored by (Starkweather, 1988) which discard the interior samples of the sample lattice. This retains empirical data in regions where the print function is most apt to be non-linear. However, it does not guarantee even $C^0$ continuity of the print function.

# References

ACM (1979). Status report of the graphics standards committee. *Computer Graphics*, 13(3). (The HLS colour space and its RGB transformations).

Acton, F. S. (1970). *Numerical Methods that Work*. Harper & Row.

Ahmed, N., Natarajan, T., and Rao, K. R. (1974). The discrete cosine transform. *IEEE Transactions on Computers*, 23(1):90–93.

Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1983). *Data structures and algorithms*. Addison-Wesley, Reading, MA.

Albert, A. (1972). *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York.

Anderson, E., Bai, Z., Bischof, C., Demmel, J., Dongarra, J., Croz, J. D., Greenbaum, A., Hammarling, S., McKenney, A., and Sorensen, D. (1991). Preliminary LAPACK users' guide. Technical report, LAPACK Project, Computer Science Department, University of Tennessee, Knoxville.

Apostol, T. M. (1969a). *Calculus*, volume 2, chapter 1.11 – Euclidean Space, page 15. John Wiley & Sons, London and New York, 2nd edition.

Apostol, T. M. (1969b). *Calculus*, volume 2, chapter 2, pages 1–2. John Wiley & Sons, London and New York, 2nd edition. (Coordinate determination).

Apostol, T. M. (1969c). *Calculus*, volume 2, chapter 5.16 – Symmetric Transformations, pages 135–138. John Wiley & Sons, London and New York, 2nd edition.

Apostol, T. M. (1969d). *Calculus*, volume 2, chapter 15.15 – Chebyshev polynomials, pages 596–598. John Wiley & Sons, London and New York, 2nd edition.

Apostol, T. M. (1969e). *Calculus*, volume 2, chapter 1.16 – Best Approximation (theorem), pages 28–30. John Wiley & Sons, London and New York, 2nd edition.

Ball, W. W. R. and Coxeter, H. S. M. (1974). *Mathematical Recreations & Essays*, pages 141–142. University of Toronto Press, 12th edition.

Barlow, H. B. (1982). What causes trichromacy? A theoretical analysis using comb-filtered spectra. *Vision Research*, 22:635–644.

Bass, D. H. (1981). Using the video lookup table for reflectivity calculations: Specific techniques and graphic results. *Computer Graphics and Image Processing*, 17(3):249–261.

Batcher, K. (1968). Sorting networks and their applications. *Conf. Proc. SJCC*, pages 307–314.

Baumgart, B. G. (1974). *Geometric Modeling for Computer Vision.* PhD thesis, Stanford University.

Birkhoff, G. and MacLane, S. (1965). *A Survery of Modern Algebra*, pages 277–278. Macmillan, New York, 3nd edition. (Section IX-14).

Blinn, J. F. (1977). Models of light reflection for computer synthesized pictures. *Computer Graphics (ACM SIGGRAPH '77 Proceedings)*, 11(2):192–198.

Blinn, J. F. (1979). Raster graphics. In Booth, K. S., editor, *Tutorial: Computer Graphics*. IEEE Computer Society.

Blinn, J. F. (1982). Light reflection functions for simulation of clouds and dusty surfaces. *Computer Graphics (ACM SIGGRAPH '82 Proceedings)*, 16(3):21–30.

Booth, K. S., Forsey, D. R., and Paeth, A. W. (1986). Hardware assistance for Z-buffer visible surface algorithms. In *Proceedings, Graphics Interface '86*, Vancouver. Canadian Information Processing Society.

Booth, K. S. and MacKay, S. A. (1982). Techniques for frame buffer animation. In *Proceedings, Graphics Interface '82*, pages 213–220.

Bouma, W. J. and Vaněček Jr., G. (1993). Modeling contacts in a physically based simulation. In *Proceedings of the Second Symposium on Solid Modeling and Applications*, pages 409–418, New York, NY. ACM Press.

Boynton, R. M. and Gordon, J. (1965). Bezold-Brücke hue shift measured by color-naming technique. *Journal of the Optical Society of America*, 55:78–86.

Bracewell, R. N. (1978). *The Fourier Transform and Its Application*, chapter 10. McGraw-Hill, New York, 2nd edition.

Brainard, D. H., Wandell, B. A., and Cowan, W. B. (1989). Black light: How sensors filter spectral variation of the illuminant. *IEEE Trans. on Biomedical Engineering*, 36:140–149.

Brill, M. H. (1978). A device performing illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology*, 71:473–478.

Brill, M. H. (1979). Further features of the illuminant-invariant trichromatic photosensor. *Journal of Theoretical Biology*, 71:305–308.

Brisson, D. W. (1978). Hypergraphics: Visualizing complex relationships in art, science and technology. In *AAAI Selected Symposium 24*, Boulder, CO. Westview Press.

Brown, W. R. J. (1955). Subtractive color reproduction: Evaluation of the actual color reproduction equations for a color process. *Journal of the Optical Society of America*, 45(7):539–546.

Buchanan, M. D. and Pendergrass, R. (1980). Digital image processing. *Electro-optical Systems Design*, 12(3):29–36.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–.

Buchsbaum, G. and Gottshalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London*, 220(B):89–113.

Buchsbaum, G. and Gottshalk, A. (1984). Chromaticity coordinates of frequency-limited functions. *Journal of the Optical Society of America*, 1(8):885–887.

Buckalew, C. and Fussell, D. (1989). Illumination networks: Fast realistic rendering with general reflectance functions. *Computer Graphics (ACM SIGGRAPH '89 Proceedings)*, 23(3):89–98.

Buckley, R., Green, D., et al. (1987). Proposal for an L*u*v* hardware frame buffer on Dorado. (private conversation, April 24th).

Burns, S. A., Cohen, J. B., and Kuznetsov, E. N. (1989). Multiple metamers: Preserving color matches under diverse illuminants. *Color Research and Application*, 14:16–22.

Cabral, B., Max, N., and Springmeyer, R. (1987). Bidirectional reflectance functions from surface bump maps. *Computer Graphics (ACM SIGGRAPH '87 Proceedings)*, 21(4):273–281.

Capelli, R. (1991). Fast approximation to the arctangent. In Arvo, J., editor, *Graphics Gems II*, chapter 8, pages 389–391. Academic Press, Boston.

Catmull, E. E. (1979). A tutorial on compensation tables. *Computer Graphics (ACM SIGGRAPH '79 Proceedings)*, 13(2):1–7.

Chan, T. F. (1986). Alternative to the SVD: Rank revealing QR-factorizations. In *Proceedings of SPIE – The International Society for Optical Engineering, V. 696. Advanced Algorithms and Architectures for Signal Processing*, pages 31–38, Bellingham, Washington. SPIE.

Chand, D. R. and Kapur, S. S. (1970). An algorithm for convex polytopes. *Communications of the ACM*, 17(1):78–86.

Chen, P.-C., Hansen, P., and Jaumard, B. (1991). On-line and off-line vertex enumeration by adjacency lists. *Oper. Res. Lett.*, 10:403–409.

Chihara, T. S. (1978). *An Introduction to Orthogonal Polynomials*. Gordon and Breach, New York.

Clapper, F. R. (1961). An empirical determination of halftone color-reproduction requirements. In *Proceedings of the Annual Technical Meeting*, volume 13, pages 31–41. Technical Association of the Graphics Arts (TAGA).

Clarke, F. J. J. and Parry, D. J. (1985). Helmholtz reciprocity: Its validity and application to reflectometry. *Lighting Research and Technology*, 17(1):1–11.

Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell colour chips. *Psychonomic Science*, 1:369–370.

Cohen, J. B. (1988). Color and color mixture: Scalar and vector fundamentals. *Color Research and Application*, 13:4–39.

Cohen, J. B. and Kappauf, W. E. (1985). Color mixture and fundamental metamers: Theory, algebra, geometry, application. *Am. Journal of Psychology*, 98:171–259.

Cohen, M. F., Greenberg, D. P., Immel, D. S., and Brock, P. J. (1986). An efficient radiosity approach for realistic image synthesis. *IEEE Computer Graphics and Applications*, 6(2):26–35.

Cohen, M. F. and Wallace, J. R. (1993). *Radiosity and Realistic Image Synthesis.* Academic Press, Cambridge, MA.

Commission Internationale D'Clairage (1933). The standard colorimetric observer. *(CIE Recommendations).*

Commission Internationale D'Clairage (1986). Colorimetry. *(CIE Recommendations)*, 15.2.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* Chapman-Hall, New York.

Cook, R. L. and Torrance, K. E. (1981). A reflectance model for computer graphics. *Computer Graphics (ACM SIGGRAPH '81 Proceedings)*, 15(3):307–316.

Cowan, W. B. (1983). An inexpensive scheme for calibration of a colour monitor in terms of CIE standard coordinates. *Computer Graphics (ACM SIGGRAPH '83 Proceedings)*, 17:315–321.

Cowan, W. B. (1986). CIE calibration of video monitors. Technical Report 2536217, National Research Council of Canada, Ottawa, Ontario.

Cowan, W. B. and Rowell, N. L. (1986). Phosphor constancy and gun independence in color video monitors. *Color Research and Application*, 11(Supplement):S34–S38.

Cowan, W. B. and Ware, C. (1985). Colour perception. In *SIGGRAPH '85 Course Notes (#3)*. Association for Computing Machinery. (Revised notes of SIGGRAPH '84 course #3).

Coxeter, H. S. M. (1961). *Introduction to Geometry*. John Wiley & Sons, New York.

Coxeter, H. S. M. (1962). The classification of zonohedra by means for projective diagrams. *Journal de Mathematiques Pures et Appliquees*, 41:137–156. (See also chapter four in *Twelve Geometric Essays*).

Coxeter, H. S. M. (1968). *Twelve Geometric Essays*. Southern Illinois University Press.

Coxeter, H. S. M. (1973). *Regular Polytopes*. Dover, New York.

Coxeter, H. S. M. (1992). Geometry using merely a compass. (Private conversation following UW/ICR Colloquium, October 2nd).

Critchlow, K. (1970). *Order in Space.* Viking Press.

Crosfield (1988). The Crosfield series of digital scanners. Crosfield Electronics, Hamelhempstead, England.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005. (Includes history and extensive bibliography).

Daubechies, I. (1992). *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, Reading and Philadelphia.

Duff, T. (1985). Compositing 3-D rendered images. *Computer Graphics (ACM SIGGRAPH '85 Proceedings)*, 19(3):41–44.

D'Zmura, M. (1992). Color constancy: Surface color from changing illumination. *Journal of the Optical Society of America*, 9A:490–493.

D'Zmura, M. and Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America*, 3(10):1622–1672.

Evans, R. M., Hanson, W. T., and Brewer, W. L. (1953). *Principles of Color Photography.* John Wiley & Sons, New York, NY.

Eves, H. (1987). Geometry (mensuration). In Beyer, W. H., editor, *CRC Standard Mathematical Tables*, chapter four. The Chemical Rubber Company (CRC) Press, Boca Raton, Fl, 28th edition.

Falk, D. S., Brill, D. R., and Stork, D. G. (1986). *Seeing the Light: Optics in Nature, Photography, Color, Vision, and Holography.* Harper and Row, New York, NY.

Faugeras, O. D. (1979). Digital color image processing within the framework of a human visual model. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-27(4).

Fedorov, E. S. (1971). *Symmetry of Crystals.* American Crystallographic Association, New York. (Translation of Симметрииа и структура кристаллов).

Feller, R. L., editor (1986). *Artists' Pigments: A Handbook of their History and Characteristics.* Cambridge University Press, London.

Ferrucci, V. (1993). Generalised extrusion of polyhedra. In *Proceedings of the Second Symposium on Solid Modeling and Applications*, pages 35–42, New York, NY. ACM Press.

Ferrucci, V. and Paoluzzi, A. (1991). Extrusion and boundary evaluation for multidimensional polyhedra. *Computer Aided Design*, 23(1):40–50.

Fishkin, K. (1990). A fast HSL-to-RGB transform. In Glassner, A., editor, *Graphics Gems*, chapter 8, pages 448–449. Academic Press.

Fiume, E. (1986). *A Mathematical Semantics and Theory of Raster Graphics.* PhD thesis, Dept. of CS, U. of Toronto.

Fiume, E. L. (1989). *The Mathematical Structure of Raster Graphics.* Academic Press, San Diego, CA.

Foley, J. D., van Dam, A., Feiner, S. K., and Hughes, J. F. (1990). *Computer Graphics: Principles and Practice*, chapter 13.3. Addison-Wesley, Reading, MA, 2nd edition. (Color Models for Raster Graphics).

Fuchs, H., Poulton, J., Paeth, A., and Bell, A. (1982). Developing pixel-planes, a smart memory-based raster graphics system. In *Conference on Advanced Research in VLSI*. Massachusetts Institute of Technology, Artech House.

Gage, editor (1985). *The Canadian Style: A Guide to Writing and Editing*, page 55 (footnote). Dundurn Press Limited, Toronto.

Gershon, R. (1987). *The Use of Color in Computational Vision*. PhD thesis, University of Toronto.

Glassner, A. (1990). *Graphics Gems*, page 326. Academic Press. ("Boxes and Spheres" section summary).

Glassner, A. S., editor (1989). *An Introduction to Ray Tracing*. Academic Press, San Diego, CA.

Golub, G. H. and Kahan, W. (1965). Calculating the singular values and pseudo-inverse of a matrix. *SIAM Journal on Numerical Analysis*, 2:205–224.

Golub, G. H. and Loan, C. F. V. (1989a). *Matrix Computations*, chapter 8.3, Computing the SVD, pages 427–436. Johns Hopkins University Press, Baltimore, Maryland, 2nd edition.

Golub, G. H. and Loan, C. F. V. (1989b). *Matrix Computations*, page 141. Johns Hopkins University Press, Baltimore, Maryland, 2nd edition. (Cholesky Factorization).

Golub, G. H. and Loan, C. F. V. (1989c). *Matrix Computations*, chapter 7.6 – Invariant Subspace Computations), pages 390–392. Johns Hopkins University Press, Baltimore, Maryland, 2nd edition. (7.6.3 - Ascertaining Jordan Block Structures).

Golub, G. H. and Loan, C. F. V. (1989d). *Matrix Computations*, chapter 3.2 – The QR Factorization, pages 211–224. Johns Hopkins University Press, Baltimore, Maryland, 2nd edition.

Golub, G. H., Luk, F. T., and Overton, M. L. (1981). A block Lanczos method for computing the singular values and corresponding singular vectors of a matrix. *ACM Transactions on Mathematical Software*, 7:149–169. (Cited in Åke Björck's bibliography on least squares, which is available by anonymous ftp from `math.liu.se` in `pub/references`).

Gordon, D. and Reynolds, R. A. (1985). Image space shading of 3-dimensional objects. *Computer Vision, Graphics, and Image Processing*, 29:361–376.

Grassman, H. (1853). Zur Theorie der Farbenmischung. *Poggendorf' Ann. Phys.*, 7(254):203–303.

Grünbaum, B. (1967). *Convex Polytopes*, chapter eighteen (arrangements of hyperplanes). John Wiley & Sons, London and New York.

Gunn, C. (1993). Discrete groups and visualization of three-dimensional manifolds. *Computer Graphics (ACM SIGGRAPH '93 Proceedings)*, 27(4):255–262.

Hall, R. and Greenberg, D. P. (1983). A testbed for realistic image synthesis. *IEEE Computer Graphics and Applications*, 3(8):10–20.

Hall, R. A. (1983). A methodology for realistic image synthesis. Master's thesis, Cornell.

Hamilton, A. G. (1989). *Linear Algebra*, page 245. Cambridge University Press, Cambridge, Great Britain.

Hanrahan, P. and Krueger, W. (1993). Reflection from layered surfaces due to subsurface scattering. *Computer Graphics (ACM SIGGRAPH '93 Proceedings)*, 27(4):165–174.

Hansen, P. C., Sekii, T., and Shibahashi, H. (1991). The modified truncated-SVD method for regularization in general form. *SIAM Journal on Scientific and Statistical Computing*, 13:1142–1150.

Heckbert, P. S. (1982). Color image quantization for frame buffer display. *Computer Graphics (ACM SIGGRAPH '82 Proceedings)*, 16(3):297–307.

Heckbert, P. S. (1985). An efficient algorithm for generating zonohedra. Technical report, NYIT Computer Graphics Lab.

Heckbert, P. S. (1991). *Simulating Global Illumination Using Adaptive Meshing*. PhD thesis, CS Division (EECS), University of California, Berkeley, CA.

Heisserman, J. and Woodbury, R. (1993). Generating languages of solid models. In *Proceedings of the Second Symposium on Solid Modeling and Applications*, pages 103–112, New York, NY. ACM Press.

Higgins, T. M. and Booth, K. S. (1986). A cel-based model for paint systems. In *Proceedings, Graphics Interface '86*, pages 82–90, Vancouver. Canadian Information Processing Society.

Higham, N. J. and Stewart, G. W. (1987). Numerical linear algebra in statistical computing. In Iserles, A. and Powell, M. J. D., editors, *The State of the Art in Numerical Analysis*, pages 41–57, Oxford. Clarendon Press.

Hill, S. (1994). Tri-linear interpolation. In Heckbert, P., editor, *Graphics Gems IV*, pages 521–525. Academic Press, Boston.

Holden, A. (1971). *Space, Shape and Symmetry.* Columbia University Press, New York.

Hong, Y. P. and Pan, C.-T. (1992). Rank-revealing QR factorizations and the singular value decomposition. *Mathematics of Computation*, 58:213–232.

Horn, A. (1950). On the singular values of a product of completely continuous operators. *Proceedings of the National Academy of Sciences*, 36:374–375.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441 and 498–520.

Hotelling, H. (1943). Some new methods in matrix calculations. *Annals of Mathematical Statistics*, 14:1–34.

Hurlbert, A. C. and Poggio, T. A. (1988). Synthesizing a color algorithm from examples. *Science*, 236:482–.

Immel, D. S., Cohen, M. F., and Greenberg, D. P. (1986). A radiosity method for non-diffuse environments. *Computer Graphics (ACM SIGGRAPH '86 Proceedings)*, 20(4):133–142.

In Der Smitten, F. J. (1974). Data-reducing source encoding of color picture signals based on chromaticity classes. *Nachrichtentech. Z.*, 27:176.

Joblove, G. and Greenberg, D. (1978). Color spaces for computer graphics. *Computer Graphics (ACM SIGGRAPH '78 Proceedings)*, 12:20–25.

Kajiya, J. T. (1985). Anisotropic reflection models. *Computer Graphics (ACM SIGGRAPH '85 Proceedings)*, 19(3):15–22.

Kajiya, J. T. (1986). The rendering equation. *Computer Graphics (ACM SIGGRAPH '86 Proceedings)*, 20(4):143–150.

Kaprass, J. (1992). *Connections: The Geometric Bridge between Art and Science*, pages 371–381. McGraw-Hill, New York, NY.

Keightley, J. (1987). Private Conversation (Pthalo Systems, Inc., Richmond, B.C.).

Koenderink, J. J. (1987). Color atlas theory. *Journal of the Optical Society of America*, 4(7):1314–1321.

Krantz, D. H. (1975). Color measurement and color theory: I. representation theorem for Grassman structures. *Journal of Mathematical Psychophysics*, 12:203–303.

Kreiger, R. (1984). 3-D environments for 2-D animation. Master's thesis, University of Waterloo, Waterloo, Ontario, Canada. (Texture-mapping in perspective).

Krinov, E. (1947). Spectral reflectance properties of natural formations. Technical Report TT-439, Canadian National Research Council, Ottawa. (Technical translation from Russian).

Lai, J. W. (1991). Implementation of colour design tools using the OSA Uniform Colour System. Master's thesis, University of Waterloo.

Land, E. H. (1977). The retinex theory of color vision. *(The) Scientific American*, pages 108–128. (See also Proc. Roy. Inst. of Great Britain v.47, 1974)).

Lansdale, R. (1991). Texture-mapping and resampling for computer graphics. Master's thesis, University of Toronto, Dept. of Elec. Eng.

Lawrie, D. (1975). Access and alignment of data in an array processor. *IEEE Trans. on Computers*, C-24(12):1145–1155.

Lawson, C. L. and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice Hall, Englewood Cliffs, New Jersey.

Leon, S. J. (1990a). *Linear Algebra with Applications*, chapter 5, page 211. Macmillan, New York, NY, 3rd edition. (Theorem 5.5.2).

Leon, S. J. (1990b). *Linear Algebra with Applications*, chapter 6, page 307. Macmillan, New York, NY, 3rd edition. (Theorem 6.5.1).

Li, X. and Strahler, A. H. (1985). Geometric-optical modeling of a conifer forest canopy. In *IEEE Trans. Geosci. Remote Sensing*, volume GE-23, pages 705–721.

Libbey, R. L. (1991). *Handbook of Circuit Mathematics for Technical Engineers*. The Chemical Rubber Company (CRC) Press, Boca Raton, Fl.

Loeb, A. L. (1970). A systematic survey of cubic crystal structures. *J. Solid State Chem.*, 1:237–267.

Loeb, A. L. (1976). *Space Structures, their Harmony and Counterpoint*. Addison-Wesley, Reading, MA.

Ludman, J. E. and Sampson, J. L. (1981). A technique for generating gray codes. *Journal of Stat. Plan Infer.*, 5:171–180.

MacAdam, D. L. (1956). Chromatic adaptation. *Journal of the Optical Society of America*, 46:500–.

MacDonald-Dettwiler and Associates (1988). The FIRE 100 film recorder. Richmond, B.C. (Proprietary internal colour correction hardware supplied by Polaroid, Inc.).

Maloney, L. (1985a). *A Computational Approach to Color Constancy.* PhD thesis, Stanford University.

Maloney, L. T. (1985b). Computational approaches to color constancy. Technical Report 1985-01, Stanford University.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with a small number of parameters. *Journal of the Optical Society of America*, 3:1673–1683.

Maloney, L. T. and Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3:29–33.

Marimont, D. H. and Wandell, B. A. (1992). Linear models of surface and illuminant spectra. *Journal of the Optical Society of America*, 9(11):1905–1913.

Martindale, D. and Paeth, A. W. (1991). Television colour encoding and "hot" broadcast colors. In Arvo, J., editor, *Graphics Gems II*, chapter 3.6, pages 147–158. Academic Press.

Matheiss, T. H. and Rubin, D. S. (1980). A survey and comparison of methods for finding all vertices of convex polyhedral sets. *Math. Oper. Res.*, 5:167–185.

McIlwain, K. and Dean, C. E. (1956). *Principles of Color Television.* John Wiley & Sons.

McKenna, M. and Seidel, R. (1985). Finding the optimal shadows of a convex polytope. In *Proc. of the Symp. on Computational Geometry*, pages 24–28.

Meyer, G. W. (1986). *Color Calculations For and Perceptual Assessment of Computer Graphic Images*. PhD thesis, Program of Computer Graphics, Cornell University, Ithica, NY.

Meyer, G. W. (1988). Color issues in synthetic image generation. In *SIGGRAPH '88 Course Notes (#12 – Guide to Image Synthesis)*, pages 49–80. Association for Computing Machinery.

Meyer, G. W. and Greenberg, D. P. (1986). Color education and color synthesis in computer graphics. *Color Research and Application*, 11(Supplement):S39–S44.

Mitchell, E. N. (1984). *Photographic Science*. John Wiley & Sons, New York, NY.

Miyazaki, K. (1986). *An Adventure in Multidimensional Space - The Art and Geometry of Polygons, Polyhedra, and Polytopes*. John Wiley & Sons, New York.

Moon, P. (1945). Polynomial approximation of reflectance curves. *Journal of the Optical Society of America*, 35.

Morris, J. L. (1983). *Computational Methods in Elementary Numerical Analysis*. Chichester Wiley, West Sussex Toronto.

Munsell, A. H. (1936-1963). *A Color Notation*. Munsell Color Company, Baltimore, MD.

Murray-Coleman, J. F. and Smith, A. M. (1990). The automated measurement of BRDFs and their application to luminaire modeling. *Journal of the Illumination Engineering Society*, 19(1):87–99.

Nassau, K. (1980). The causes of color. *(The) Scientific American*, 219(10):124–154.

Neugebauer, H. E. J. (1937). Die Theoretishchen Grundlagen des Mehrfarbenbuchdrucks. *Z. Techno. Phys.*, 36:75–89.

Neuman, L. and Neumann, A. (1989). Photosimulation: Interreflection with arbitrary reflectance models and illumination. *Computer Graphics Forum*, 8:21–34.

Newman, M. and Sproull, R. F. (1971). *Introduction to Computer Graphics*, chapter 12 - Three Dimensional Transformations and Perspectives, pages 247–249. Addison-Wesley, 1st edition.

Nicodemus, F. E., editor (1978). *Self-Study Manual on Optical Radiation Measurements*. NBS-1,2,3. National Bureau of Standards, Washington, DC.

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T. (1977). Geometrical considerations and nomenclature for reflectance. Technical report, National Bureau of Standards, US Dept. of Commerce, Washington, DC.

Noll, A. M. (1967). A computer technique for displaying n-dimensional hyperobjects. *Communications of the ACM*, 10(8):469–473.

Novak, C. C. and Shafer, S. A. (1990). Supervised color constancy using a color chart. Technical Report CMU-CS-90-140, University of Buffalo at SUNY.

Overton, M. C. (1990). Optimization of eigenvalues. (Private conversation following UW/ICR Colloquium, March 14).

Paeth, A. W. (1986). Design and experience with a generalized raster toolkit. In *Proceedings, Graphics Interface '86*, pages 77–81, Vancouver. Canadian Information Processing Society.

Paeth, A. W. (1987a). The IM raster toolkit – design, implementation, and use. Technical Report ICR 87-03, University of Waterloo.

Paeth, A. W. (1987b). Theoretical models for digital color. (Second Stage Thesis Examination, September 15).

Paeth, A. W. (1988). The Lemming editor. IRIS Software Exchange 17(1), Silicon Graphics, Inc.

Paeth, A. W. (1989a). Algorithms for fast color correction. In *Int. Symp. Digest of Technical Papers*, volume 30, pages 169–175. Society for Information Display (SID).

Paeth, A. W. (1989b). Fast algorithms for color correction. Technical Report CS-89-42, University of Waterloo.

Paeth, A. W. (1990a). Mapping RGB triples onto four bits. In Glassner, A., editor, *Graphics Gems*, chapter 4, pages 233–245. Academic Press.

Paeth, A. W. (1990b). Proper treatment of pixels as integers. In Glassner, A., editor, *Graphics Gems*, chapter 4, pages 249–258. Academic Press.

Paeth, A. W. (1990c). Proper treatment of pixels as integers. In Glassner, A., editor, *Graphics Gems*, chapter 4, pages 252–3. Academic Press. (See also *ACM Transactions on Graphics* **6**(3) plates 15-18 (p. 235)).

Paeth, A. W. (1991). Mapping RGB triples onto 16 distinct values. In Arvo, J., editor, *Graphics Gems II*, chapter 3.5, pages 143–146. Academic Press.

Paeth, A. W. (1992). A generic pixel selection mechanism. In Kirk, D., editor, *Graphics Gems III*, chapter 2.6, pages 77–79. Academic Press.

Paige, C. C. (1980). Error analysis of some techniques for updating orthogonal decompositions. *Mathematics of Computation*, 34:465–471.

Panter, P. F. (1965). *Modulation, Noise, and Spectral Analysis*. McGraw-Hill.

Parkkinen, J. P. S., Hallikainen, J., and Jaaskelainen, T. (1989). Characteristic spectra of Munsell colour. *Journal of the Optical Society of America*, 6:318–322.

Peachey, D. R. (1987). Pixel-buffer visible surface algorithms. In *Workshop on Rendering Algorithms and Systems, CHI+GI '87*. Canadian Information Processing Society, Toronto.

Pegna, J. (1987). *Variable Sweep Geometric Modeling*. PhD thesis, Stanford University, Stanford, CA.

Penna, A. and Patterson, R. (1986). *Projective Geometry and Its Applications to Computer Graphics*. Prentice-Hall.

Phong, B.-T. (1975). Illumination for computer generated images. *Communications of the ACM*, 18(6):311–317.

Pique, M. E. (1983). Fast 3-D display of space-filling molecular models. Technical Report CS-83-004, U. of North Carolina, Chapel Hill, N.C.

Pobboravsky, I. (1966). Methods of computing ink amounts to produce a scale of neutrals for photomechanical reproduction. In *Proceedings of the Annual Technical Meeting*. Technical Association of the Graphics Arts (TAGA).

Pobboravsky, I. and Pearson, M. L. (1972). Computation of dot areas required to match a colorimetrically specified color using the modified neugebauer equations. In *Proceedings of the Annual Technical Meeting*. Technical Association of the Graphics Arts (TAGA).

Pollak (1955). Masking for halftone. *Journal of Photographic Science*, 3:180–188.

Poulin, P. and Fournier, A. (1990). A model for anisotropic reflection. *Computer Graphics (ACM SIGGRAPH '90 Proceedings)*, 24(3):273–282.

Poulton, J., Fuchs, H., and Paeth, A. (1985). Pixel-planes graphic engine (case study). In Weste, N. and Eshraghian, K., editors, *Principles of CMOS VLSI Design*. Addison-Wesley, Reading, MA.

Preparata, F. (1977). Convex hulls of finite sets of points in two and three dimensions. *Communications of the ACM*, 20(7):87–93.

Preparata, F. (1979). An optimal real-time algorithm for planar convex hulls. *Communications of the ACM*, 22(7):402–405.

Preparata, F. P. and Shamos, M. I. (1985). *Computational Geometry: an Introduction*. Springer-Verlag, New York, NY.

Press, W. H. et al. (1992). *Numerical Recipes in C: the Art of Scientific Computing*. Cambridge University Press, Cambridge, England, 2nd edition.

Pringsheim, P. (1949). *Fluorescence and Phosphorescence*. Interscience, New York.

Pulleyblank, W. R. (1989). A survey and comparison of methods for finding all vertices of convex polyhedral sets. *Math. Oper. Res.*, 5:167–185.

Purdy, D. M. (1931). Spectral hue as a function of intensity. *American Journal of Psychology*, 43:541–.

Purkinje, J. (1823). *Beobachtungen und Versuche zur Physiologie der Sinne.* G. Reiner, Berlin, (Zweites Bändchen) 2nd edition.

Raso, M. (1990). A piecewise polynomial approach to shading using spectral distributions. Master's thesis, Dept. of CS, University of Toronto.

Raso, M. and Fournier, A. (1991). A piecewise polynomial approach to shading using spectral distributions. In *Proceedings, Graphics Interface '91*, pages 82–90, Calgary, AB. Canadian Information Processing Society.

Ratliff, F. (1965). *Mach Bands: Quantitative Studies on Neural Networks in the Retina.* Holden-Day, San Francisco.

Rawlins, G. J. E. (1992). *Compared to what? : An Introduction to the Analysis of Algorithms*, page 178. Computer Science Press, New York. (On algorithm construction using an adversary).

Rice, J. R. (1983). *Numerical methods, software, and analysis.* McGraw-Hill, New York.

Robertson, A. R. (1977). The CIE 1976 color-difference formulae. *Color Research and Application*, 2:7–11.

Rudin, W. (1973). *Functional Analysis*, chapter 7. Fourier Transforms. McGraw-Hill, New York. (normalized Lesbesque measures).

Ruhe, A. (1983). Numerical aspects of Gram-Schmidt orthogonalization of vectors. *Linear Algebra and Its Applications*, 53/53:591–601.

Rushmeier, H. E. and Torrance, K. E. (1990). Extending the radiosity method to include specularly reflecting and translucent materials. *ACM Transaction on Graphics*, 9(1):1–27.

Sällström, P. (1973). Colour and physics: Some remarks concerning the physical aspects of human colour vision. Technical Report 73-09, Institute of Physics, University of Stockholm.

Santoro, N. (1984). Chain multiplication of matrices of approximately or exactly the same size. *Communications of the ACM*, 27(2):152–156.

Schwarz, M. W., Cowan, W. B., and Beatty, J. C. (1987). An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics*, pages 123–158.

Scott, D. S. (1979). Block Lanczos software for symmetric eigenvalue problems. Technical Report 84-08, Oak Ridge National Laboratory.

Seidel, R. (1981). A convex hull algorithm optimal for point sets in even dimensions. Technical report, Dept. of CS, University of British Columbia, Vancouver, B.C.

Seidel, R. (1990). Linear programming and convex hulls made easy. In *Proc. 6th Ann. Sympos. Computatoinal Geometry*, volume 5, pages 167–185.

Shafer, S. (1982). Describing light mixtures through linear algebra. *Journal of the Optical Society of America*, 72:299–300.

Shafer, S. (1985). Using color to separate reflection components. *Color Research and Application*, 10:210–218. (See also University of Rochester report CS-TR-136).

Shoup, R. G. (1979). Color table animation. *Computer Graphics (ACM SIGGRAPH '79 Proceedings)*, 13(2):8–13.

Sillion, F., Arvo, J., Westin, S., and Greenberg, D. (1991). A global illumination algorithm for general reflection distributions. *Computer Graphics (ACM SIGGRAPH '91 Proceedings)*, 25(4):187–196.

Sillion, F. and Puech, C. (1989). A general two-pass method integrating specular and diffuse reflection. *Computer Graphics (ACM SIGGRAPH '89 Proceedings)*, 23(3):335–344.

Smallman, H. S. and Boynton, R. M. (1990). Segregation of basic colors in an information display. *Journal of the Optical Society of America*, 7(A)(10):1985–1994.

Smith, A. R. (1976). RGB to HSV transforms. Technical report, NYIT Computer Graphics Lab.

Smith, A. R. (1978a). Color gamut transform pairs. *Computer Graphics (ACM SIGGRAPH '78 Proceedings)*, 12(3):12–19.

Smith, A. R. (1978b). Realizable colors. Technical report, NYIT Computer Graphics Lab.

Smith, A. R. (1979). YIQ vs. RGB. Technical report, NYIT Computer Graphics Lab.

Smits, B. E. and Meyer, G. W. (1990). Newton's colors: Simulating interference phenomena in realistic image synthesis. In *Proceedings of the Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics*, pages 185–194.

Spencer, D. E. and Gaston, E. G. (1975). Current definitions of reflectance. *J. Opt. Soc. Amer.*, 65(10):1129–1132.

Stahl, H. and Totik, V. (1992). *General Orthogonal Polynomials*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, England.

Starkweather, G. (1988). Continuous color correction. *Annual Meeting Digest, Inter-Society Color Council/SID.*

Stenger, L. (1977). Quantization of TV chrominance signals considering the visibility of small color differences. *IEEE Trans. on Communications*, COM-25(11):1393.

Stewart, G. W. (1973). *Introduction to Matrix Computations*. Academic Press, New York.

Stewart, G. W. (1979). A note on the perturbation of singular values. *Linear Algebra and Its Applications*, 28:213–216.

Stewart, G. W. (1984). A second order perturbation expansion for small singular values. *Linear Algebra and Its Applications*, 56:231–235.

Stiles, W. S. (1961). Adaptation, chromatic adaptation, colour transformation. *Anales Real Soc. Espan. Fis. Quim.*, 57A:149–175.

Stiles, W. S., Wyszecki, G., and Ohta, N. (1977). Counting metameric object-color stimuli using frequency-limited spectral reflectance functions. *Journal of the Optical Society of America*, 67:779–.

Stone, M. C., Beatty, J. C., and Cowan, W. B. (1986). A description of the color-reproduction methods used for this issue of colour research and application. *Color Research and Application*, 11(Supplement):S83,S88.

Stone, M. C., Cowan, W. B., and Beatty, J. C. (1988). Color gamut mapping and the printing of digital color images. *ACM Trans. on Graphics*, 7(4):249–292.

Stone, M. C. and Wallace, W. E. (1991). Gamut mapping computer generated imagery. In *Proceedings, Graphics Interface '91*, pages 32–39.

Strang, G. (1989). Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31(4):614–627.

Sugihara, K. (1993). Resolvable representation of polyhedra. In *Proceedings of the Second Symposium on Solid Modeling and Applications*, pages 127–135, New York, NY. ACM Press.

Takahama, K. and Nayatani, Y. (1972). New method for generating metameric stimuli of object colors. *Journal of the Optical Society of America*, 62:1516–1520.

Tansley, B. W. and Boynton, R. M. (1977). Metric of color borders. *Science*, 197:1196.

Tatsuoka, M. M. (1988). *Multivariate Analysis: Techniques for Educational and Psychological Research.* Macmillan, New York, 2nd edition.

Taylor, E. S. (1974a). *Dimensional Analysis for Engineers.* Claridon Press, Oxford.

Taylor, M. M. (1973). Principal components colour display of ERTS imagery. In *Proc. Third Earth Resources Technology Satellite Symp.*, volume NASA SP-351, pages 1877–1897.

Taylor, M. M. (1974b). Principal components colour display of ERTS imagery. In *Proc. Second Canadian Symp. Remote Sensing*, pages 295–313.

Tominaga, S. and Wandell, B. A. (1989). Standard surface-reflectance model and illuminant estimation. *Journal of the Optical Society of America*, 6(4):576–584.

Tominaga, S. and Wandell, B. A. (1990). Component estimation of surface spectral reflectance. *Journal of the Optical Society of America*, 7(2):312–317.

Torrance, K. E. and Sparrow, E. M. (1967). Theory of off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114.

Turkowski, K. (1986). Anti-aliasing in topological color spaces. *Computer Graphics (ACM SIGGRAPH '86 Proceedings)*, 20(4).

von Kries, J. (1904). Die Gesichtsempfindungen. In *Handbuch der Physiologie der Menchen*, volume 3, pages 211–. Nagel.

Wandell, B. A. (1989). Color constancy and the natural image. *Physica Scripta*, 39:187–192.

Watkins, D. S. (1991a). *Fundamentals of matrix computations*, chapter 3.2 – Orthogonal Matrices, Rotators, and Reflectors, page 132. John Wiley & Sons.

Watkins, D. S. (1991b). *Fundamentals of matrix computations*, chapter 7.5 – Angles and Distances Between Subspaces, pages 421–430. John Wiley & Sons.

Watkins, D. S. (1991c). *Fundamentals of matrix computations*, chapter 4.3 – The Power Method and Some Simple Extensions, pages 210–224. John Wiley & Sons.

Weld, J. (1987). *Geometric Representation of Swept Volumes with Application to Polyhedral Objects.* PhD thesis, Cornell University.

Wesolowsky, G. O. (1976). *Multiple Regression and Analysis of Variance.* John Wiley & Sons, New York.

Whitted, T. (1980). An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349.

Whitten, M. and England, N. (1987). The TAAC hardware accelerator. (Private Conversation with Transcept, Inc.).

Wilkinson, J. H. (1965). Error analysis of transformations based on the use of matrices of the form $I-2xx^{\mathrm{H}}$. In Rall, L. B., editor, *Error in Digital Computation*, pages 77–101. John Wiley, New York.

Wilkinson, J. H. and Reinsch, C. (1971). *Handbook for Automatic Computation. Vol. II Linear Algebra.* Springer, New York.

Wright, W. D. (1934). The measurement and analysis of colour adaptation phenomena. *Proceedings of the Royal Society*, 115B:49–.

Wyszecki, G. (1958). Evaluation of metameric colors. *Journal of the Optical Society of America*, 48:451–454.

Wyszecki, G. (1981). Color appearance. In Kenneth R. Boff, L. K. and Thomas, J., editors, *Visual Sensory Processes and Perception*, volume 1 of *Handbook of Perception and Human Performance*, chapter 9. Wiley, Toronto and New York.

Wyszecki, G. and Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley & Sons, London and New York, 2nd edition.

Wyszecki, G. and Styles, W. S. (1982a). *Color Science : Concepts and Methods, Quantitative Data and Formulae*, chapter 3.8.5, page 208. John Wiley & Sons, London and New York, 2nd edition. (Band-limited Reflectances).

Wyszecki, G. and Styles, W. S. (1982b). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, chapter 3.2.2 – Trisimulus Space, page 119. John Wiley & Sons, London and New York, 2nd edition.

Wyszecki, G. and Styles, W. S. (1982c). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, chapter 3.2.4 – Imaginary Color Stimuli, page 127. John Wiley & Sons, London and New York, 2nd edition.

Wyszecki, G. and Styles, W. S. (1982d). *Color Science: Concepts and Methods, Quantitative Data and Formulae*, chapter 3.8.2 – Metameric Black, pages 187–188,781. John Wiley & Sons, London and New York, 2nd edition.

Yilmaz, H. (1962). Color vision and a new approach to general perception. In Bernard, E. E. and Kare, M. R., editors, *Biological Prototypes and Synthetic Systems*. Plenum, New York.

Young, R. A. (1986). Principle-component analysis of macaque lateral geniculate nucleus chromatic data. *Journal of the Optical Society of America*, 3A:1735–1742.

Yule, J. A. C. (1967a). *Principles of Color Reproduction*, pages 312–313. John Wiley & Sons, London and New York.

Yule, J. A. C. (1967b). *Principles of Color Reproduction (applied to photomechanical reproduction, color photography, and the ink, paper, and other related industries)*. John Wiley & Sons, London and New York.

Yule, J. A. C. and Claper, F. R. (1960). Undercolor removal requirements of the black printer. *The Graphics Arts Monthly*. (First in series).

Zavada, R. J. (1988). Challenges to the development of a standardized professional studio color-picture monitor. *SMPTE Journal*, pages 703–710.