Abstract— The interaction of light and object surfaces generates color signals in the visible band that are responsible for digital acquisition systems outputs. Inverting this mapping from the sensor space back to the wavelength domain is of great interest for many applications. Since 1964, with the idea of Jozef Cohen to exploit the characteristic of smoothness of surface reflectance functions, a lot of work has been done in the analysis, synthesis and recovering of spectral information using linear models. The general use of such models is for the establishment of a one-to-one relationship between sensor’s data and reflectance spectrum, with the requirement of ensuring the quality of the recovered spectrum in terms of physical feasibility and naturalness. In this paper we propose a solution to correct the outcome of a generic recovery method, in order to take into account quality constrains. Our strategy assumes the smoothness of the solution of the recovery method, an assumption implicitly satisfied from the adoption of linear models to represent reflectance functions.

Index Terms—

I. INTRODUCTION

The color of objects may be communicated in many forms. Color may be described by samples, by color terms or by numerical parameters that encode its appearance. The description that completely characterizes the physical property of surfaces responsible for color is the reflectance spectrum, a function defined on the domain of visible wavelengths that represents the percentage of incident light that the surface reflects at any wavelength. The product of the surface reflectance and the spectral power distribution of the illuminant defines, for given incident and viewing geometries, the color signal [Buchsbaum1984], which, entering the eye is filtered by the photoreceptors to determine at post receptorial and cognitive levels the perceived color. Likewise, the color signal is filtered and processes in digital cameras to give a color description in the device color space. The problem of relating measurements like those coming from a color measuring device, with spectral information—often in the form of a reflectance function—has many applications, such as estimating a reflectance function given the outputs of a camera system for the characterization of such devices [Cheung2005]; or inverting a spectrum-to-colorimetry mapping for a given color reproduction process for the simulation of the behavior of imaging systems [Sharma2002]. Typical examples of the recovery of spectral data for color surfaces is their rendering under different illuminants, a procedure known as color correction. Color correction methods simulate different illuminants by computing tristimulus values under a target illuminant, given the tristimulus values under a reference illuminant. Color correction may exploit the knowledge of the reference illuminant to recover a reflectance spectrum to increase the simulation accuracy [Vrhel1992][Zuffi2005]. Spectral recovery, i.e. the estimation of the spectra of objects surfaces from data having lower dimensionality (i.e. from the RGB values of an image, from LMS cone signals or from a generic N-filters camera responses) is based on the limitation in bandwidth, in terms of Fourier frequency, of the reflectance spectra of the majority of natural surfaces, which are, for most natural objects, smooth and slowly varying functions of the wavelength, just like are the spectra produced in photography, printing processes, or painting [Sharma1997]. There are exceptions, such as the surface reflectance spectra of some earth metals and some animals (e.g. wings of insects, fish scales and birds feathers), but in general, the overwhelming majority of functions spectra are smooth and slowly varying functions of the wavelength in the visible range [Stiles1962][Stiles1977]. Since the work of Stiles et al. [Stiles1977], who observed the smooth spectral profiles of color signals, it has been widely demonstrated that the reflectance spectrum of natural surfaces is a smooth, low-pass function of wavelength. Band-limited functions can be adequately approximated with linear models having a small number of basis functions. As a consequence of the sampling theorem, the frequency content in the Fourier domain of a signal determines the minimum number of parameters of its linear model: if the number of parameters is limited to three, the corresponding frequency limit is approximately 0.005 cycle/nm [Buchsbaum1984]. A study of Maloney [Maloney1986] investigated the frequency content of the reflectance functions in terms of Fourier analysis of natural and artificial surfaces (the Nickerson-Munsell set and natural formations collected by Krinov [Krinov1947]), and related it to the number of basis functions necessary for their
characterization. Values of 0.01 and 0.015 cycle/nm were reported as band limits of surface reflectance functions, which, consistently with the results of Stiles [Stiles1977] of 0.01-0.02 cycle/nm, correspond to a linear model with 6 to 12 parameters. A similar frequency cut-off was observed more recently on a study conducted on a more general dataset, including natural surfaces of fruits, flowers and leaves [Bonnardel2000]. As already said, the color signal is the product of the reflectance function and the illuminant spectral power distribution: an upper bound for its frequency limit can be therefore evaluated by the convolution theorem in the sum of the frequency limits of the reflectance and illuminant functions. An analysis conducted by Romero et al. [Romero2003] considered the frequency limit of color signals, corresponding to biochrome and nonbiochrome surfaces illuminated by daylight, incandescent and fluorescent illuminants. In these signals, as far as daylight and incandescent illuminants are concerned, a limit of 0.016 cycle/nm can be considered. Fluorescent illuminants limit the possibility to adopt models with a small number of parameters due to the information content at high frequencies. On the other hand, daylights have a very low frequency limit, suggested in 0.0033 cycle/nm [Bonnardel2000] and can be modeled with very few parameters: three components are sufficient [Judd1964][Romero2003].

In the recovery of the reflectance spectra, one may assume that the illuminant is unknown, and, considering a suitable basis set, model it with a linear model as well [Wandell1987][Cheng1998]. More frequently, one assumes that the illuminant of the scene is known or can be previously estimated, and the measurements are sensors values at each point of the image. The problem therefore entails inverting, at each pixel, the sensor equations (possibly with some spatial constraints). This constitutes an under-determined, constrained problem: under-determined because, in general, for a given illuminant there will be several (uncountably many, in fact) metameric functions that, under that illuminant, will produce the desired sensor response; constrained because one seeks solutions that are physically feasible, a constraint that entails, inter alia, that the values of the reconstructed spectra fall into a the range [0,1]. This constraint is due to the fact that physically a reflectance spectrum is defined as the fraction of the incident light that the body reflects at each wavelength. Most of the standard methods for inverting the sensor’s equations do not guarantee the feasibility of the solution they find. Feasible solutions can be found by associating a cost to the sensor’s equations (a cost that would be zero when the equations are satisfied), and solving a non-linear constrained optimization problem [Schettini1996] or, considering the equations as a linear constraint, a linear programming problem [Press1986]. These methods, however, are often computationally cumbersome and can have additional drawbacks. In the case of non-linear optimization, the method may fail to find a solution even if one existed while in the case of linear programming (a method that we will consider more extensively in the following, as it will constitute part of our numerical analysis) one has little control over the properties of the solution that is found.

There is a considerable research literature on methods for spectral recovery. The analysis in the frequency domain of reflectance spectra and color signals motivates the recourse to dimensionality reduction techniques involving, in the most common approach, empirical linear models. In general, when representative data are available, linear models are defined on the basis of statistical information, applying Principal Component Analysis (PCA) [Tzeng2005] and Independent Component Analysis (ICA) [Hyvarinen2000]. The number of basis functions necessary to accurately represent the reflectance spectra depends on the characteristics of the data that one is modeling, and on the characteristics of the functions used in the linear model. There are many studies on the dimensionality of such linear models based on PCA. PCA seeks the set of basis functions that minimizes the correlation among dimensions and identifies those dimensions that are most descriptive of the data set [Wandell1987]. In general, for natural reflectance spectra, a number of basis functions between 6 and 9 is considered adequate, as emerged from studies from Cohen [Cohen1964] on a subset of the Munsell surface spectral reflectances collected by Kelley et al.[Kelley1943], from Maloney [Maloney1986] on spectral reflectances of natural formations collected by Kirnov [Kirnov1947], and from Jaaskelainen [Jaaskelainen1990] on samples derived from several different kinds of plants. Similar studies on skin reflectance indicate that three functions are sufficient [Sun2001]. A comparison of the linear model representation of the Munsell color chips with three basis components can be found in [Ramanath2004]. The methods considered, PCA, ICA, and Non-negative Matrix Factorization [Lee1999][Lee2001], performed similarly. Linear models based on PCA components or on functions computed with ICA assume a knowledge in advance of the surfaces to model. More general representations adopt generic basis functions, such as Fourier functions. In an early study from Wandell [Wandell1987], the surface reflectances of a set of 462 Munsell chips have been modeled with a three-dimensional linear system composed of Fourier basis functions, with a rather good fit between measured spectra and linear model representations. Many of the studies on the dimensionality of linear models evaluate a degree of fit between the reflectance function and its representation in the wavelength domain. But in spectral information recovery from low-dimensional data, the definition of the linear system based on a least squares cost function on reflectance representation may be inappropriate [Marimont1992]. In fact a least-square criterion ignores potential effects of the sensor sensitivities in the model fit. In selecting linear models for human vision, a weighted least-squares criterion across wavelength may be considered. According to Maloney, in fact, surface spectral reflectances fall within a linear model composed by five to seven parameters (basis reflectance vectors), but when the effect of human photoreceptors sensitivities is included, linear model with as few as three to four parameters provide...
excellent fit to the data sets [Maloney1986]. Linear models based on the L$_2$ metric are optimal when the input data follow a Gaussian distribution. When the data deviate from normal, a non-linear approach may improve linear estimation methods [DiCarlo2003]. When it is not possible to exploit previous knowledge on data, one can use linear models with generic function components [Wandell1987][Drew1992][Zuffi2004], or exploit different strategies that do not assume any a priori knowledge. A comparison of many of these methods to recover a reflectance spectrum of a set of textile samples from colorimetric triplets was reported by Dupont [Dupont2002]. Dupont considered the simplex method, the simulated annealing method, the Hawkyard method [Hawkyard1993], genetic algorithms, and neural networks. The simplex method set to assign a reflectance value for each wavelength gave very chaotic results, physically unacceptable. To solve this problem, the simplex was used to assign the proportion (between 0 and 1) of a bell curve for each wavelength. For accuracy, the simplex and Hawkyard methods were superior to the others, and the simplex coupled with the bell functions gave the most realistic shape of the curves. The high computational cost of these methods, however, was a drawback. The Hawkyard method has been recently modified to reduce computation time [Wang2005]. The methods were evaluated according to colorimetric differences, but no information about spectral match between the real and the recovered spectra was provided. The approaches investigated by Dupont [Dupont2002] were not based on training datasets. A further method that does not require the knowledge of statistical information about the reflectance spectra of the input material was proposed by Li and Luo [Li2001]. Their work was motivated by that of van Trigt [vanTrigt1990a][vanTrigt1990b], and uses a smoothness condition for recovering the reflectance of a set of tristimulus values, given the illuminant. The smoothness constraint corresponds to the integral of the first derivative of the reflectance curve on the whole wavelength band, and the feasibility of the solution is imposed in the optimization.

In recovery methods based on training sets, the occurrence of negative values in spectral recovery may indicate that the input color is out of the gamut defined by the recovery model. For this reason, in many of the methods based on statistical analysis of input data, the feasibility problem is not addressed, or non-feasible solution are omitted [Drew1992]. A method to correct reflectance functions that have negative values based on metameric black functions is described in Wyszecki and Stiles ([Wyszecki1982] pag.187). A metameric black is a spectrum that, for a specific illuminant, produces tristimulus values equal to zero. For any given illuminant, there are infinite metameric blacks, which have the feature of being positive at some wavelength and negative at others. The procedure is to add a metameric black to a given spectral reflectance function. To ensure that the resulting curve is positive for all wavelengths, the metameric black function is multiplied by an appropriate scaling factor. The metameric black approach was exploited by Morovic et al. [Morovic2005] to define the infinite set of metameric solutions obtainable by adding to the solution of an RGB-to-spectrum problem an arbitrarily scaled metameric black. To get the solution to the recovery problem, an appropriate selection within the metameric set is performed, in search of a function satisfying the properties of: physical realisability (the belonging to the range [0,1], a constraint that we call feasibility), smoothness (satisfied by linear model approach) and naturalness (property of the solution to be realizable as a convex combination of existing surface reflectances).

In this paper we present a method of a different nature to find a feasible reflectance spectrum. Our technique consists in using a simple linear method to find a (possibly unfeasible) reflectance spectrum that generates the required sensor measurements, and then to apply repeatedly a transformation that, while maintaining metamerism, will make the spectrum converge towards a feasible one. The method we propose may contribute to render applicable recovery methods based on linear system solutions that do not address the problem of the feasibility while computing the reflectance spectrum. The method we propose is based on the assumption that the recovery procedure computes a smooth function. As previously discussed, many of the methods found in the literature are of this kind, as the smoothness of the reflectance function is the primary assumption made to estimate a reflectance function at many wavelengths points from a smaller set of sensor’s responses.

II. THE MODEL

In a simplified model of light-surface interaction, a single light source is absorbed and reemitted by just one surface, producing a color signal that impinges on the retina or on the acquisition sensor. The effect that the color signal would generate on the human eye or on the imaging device can be expressed in a model where human photoreceptors or acquisition sensor can be equivalently described by their sensitivity functions. The same model, if sensitivity functions are replaced by the color matching functions, holds for the calculus of tristimulus values, as long as normalization factors are not considered. In our work, we consider the tristimulus equations as our light-surface-sensor interaction model. The tristimulus values equations are

$$X^k = \int_{\lambda} I(\lambda)x^k(\lambda)r(\lambda)d\lambda$$

where $I(\lambda)$ is the spectral power distribution of the illuminant, and $r(\lambda)$ is the surface reflectance function. Our goal is to solve the above equations for the unknown function $r(\lambda)$. Note that, for the sake of compactness, we use the letter $\Lambda$ to denote the visible wavelength interval (in this paper we always use $\Lambda=[\lambda_{\min}, \lambda_{\max}]=[400nm, 700nm]$), and write the tristimulus values as $X = [X^1, X^2, X^3]$ rather than in the customary way as $[X, Y, Z]$; correspondingly, the color matching functions are written as $x(\lambda) = [x^1(\lambda), x^2(\lambda), x^3(\lambda)]$. Both $X$ and $x$ are column vectors: throughout the paper, apices
will denote row numbers and pedices will denote column numbers. The tri-stimulus equations are typically used in their sampled form. For \( N > 2 \), \( \Delta = (\lambda_1 \ldots \lambda_N) / (N-1) \), let \( \lambda^i = \lambda_0 + (i-1) \Delta \), \( i = 1 \ldots N \); the tri-stimulus equation can then be approximated as:

\[
\sum_{i=1}^{N} I(\lambda^i) x^k(\lambda^i) r(\lambda^i) \Delta = X^k
\]

(2)

Defining \( r = [r^i] = [r(\lambda^i)] \), \( \hat{J} = \text{diag}(I(\lambda^i)) \), \( S = [x^k(\lambda^i)] = [x^k(\lambda^i)] \), we obtain the matrix form of the tri-stimulus equations: \( S = \hat{J} r \Delta \).

(3)

In order to simplify the notation, define the matrix \( A = S \Delta \) so that the tri-stimulus equations can be written as \( X = Ar \) or:

\[
\sum_{i=1}^{N} a^i_k r^i = X^k
\]

(4)

We express reflectances as a linear combination of basis functions: given \( b \) functions, \( \Psi_1 \ldots \Psi_b \), the reflectance \( r \) is expressed as:

\[
r(\lambda) = \sum_{b=1}^{b} w^b \Psi^b(\lambda)
\]

(5)

where \( w^b \) are suitable coefficients. Defining the matrix \( \Psi = [\Psi^b] = [\Psi^b(\lambda^i)] \), (4) can be rewritten in terms of the coefficients \( w \) as

\[
X^k = \sum_{i=1}^{N} \sum_{b=1}^{b} a^i_k \Psi^b(\lambda^i) \Delta = \sum_{i=1}^{N} \sum_{b=1}^{b} c^i_k w^b
\]

or \( X = A \Psi w = C w \), with \( C = A \Psi \).

In order to estimate the reflectance function given the tri-stimulus values and the spectrum of the illuminant, one must solve (6) for the coefficients \( w \), and then apply (5) to derive the sampled reflectance spectrum. If the number of basis functions is limited to three, then the tri-stimulus equations (6) can be solved by inverting the \( 3 \times 3 \) matrix \( A \Psi \). If the number of functions is greater than three, then standard linear algebra methods can be used to find a solution with certain optimality conditions, such as the minimal norm solution or a solution with the maximum number of zero components [Press1986]. These methods, as mentioned in the introduction, do not guarantee that the computed spectrum is feasible, that is, that all its components lie in the desired interval. One solution that has sometimes been used in practice is to clip the spectrum forcing it to the desired range. This solution, however, causes a distortion, possibly quite severe, in the reproduction of the color. Our method, as we mentioned briefly before, consists in using the linear method to obtain an initial spectrum \( r \), which might not be in the desired range, and then to apply iteratively a correction that will produce metameric spectra closer to the desired range. Note that any method that can derive a smooth spectrum from sensor’s data can be used in place of the linear method. Before presenting our strategy, however, it is opportune to spend a few words summarizing some facts about the possible ranges of the reconstructed spectra and the solutions that they can generate. The reflectance spectrum \( r \) specifies, for each wavelength \( \lambda \), the fraction of incident light that an object—in a given geometrical configuration with respect to the light sources and the observer—reflects. Being a fraction, it is natural to assume that, for each \( \lambda \), \( r(\lambda) \in [0,1] \) (we are assuming that objects do not emit light, and ignore translucency). We call a spectrum that satisfies this condition feasible in the strong sense, or strongly feasible. When solving the mathematical reconstruction problem, however, one can sometimes accept a weaker constraint. Consider a spectrum in the range \([0, M] \), with \( M > 1 \). Since the spectrum satisfies the tri-stimulus equations, we can write

\[
\int \lambda d\lambda, I(\lambda) x^k(\lambda) r(\lambda) = \int \lambda d\lambda, M I(\lambda) x^k(\lambda) \frac{r(\lambda)}{M} = X^k
\]

(7)

where, of course, \( \bar{r}(\lambda) = r(\lambda) / M \in [0,1] \) is strongly feasible. Any spectrum \( r \) in the range \([0, M] \) can be transformed in a strongly feasible spectrum \( \bar{r}(\lambda) \) and although \( \bar{r}(\lambda) \) is not, stricti dictu, a metamer of \( r(\lambda) \), the same color can be generated simply by increasing the intensity of the illuminant from \( I(\lambda) \) to \( M I(\lambda) \). For some applications this may be acceptable, leaving on the reflectance only the constraint \( r(\lambda) \geq 0 \). We call a spectrum that satisfies this condition feasible in the weak sense, or weakly feasible. Intuitively, weak feasibility is, as the name implies, a less demanding condition to impose. For a given illuminant, the feasibility conditions on \( r \) place some constraints on the tri-stimulus values \( x \) that one can obtain, that is, on the colors that can be physically realized. Assuming for the sake of simplicity weak feasibility and the discrete model (4), let \( e_i \) be the \( i \)-th element of the natural basis of the space \( R^N \) of the discrete spectra. Each spectrum \( r \) can be written as:

\[
r = \sum_{i=1}^{N} r^i e_i
\]

(8)

with \( r^i \geq 0 \). If we vary the \( b \)-th component of \( r \) and leave the rest of it to zero, that is, if we take \( r = r^b e_b \), we obtain, from the tri-stimulus equation, \( X^k = r^b a^b_k \) where \( r^b \geq 0 \) because of the feasibility constraint, and \( a^b_k \geq 0 \) because \( I \geq 0 \) and all the \( x^b \) are positive [Wyszecki1982]. The locus of the tri-stimulus \( X \) is in this case a half-line in the positive octant of the space \( X^1, X^2, X^3 \). Varying different components of \( r \) we obtain \( N \) such lines. Equation (4) states that, if the feasibility constraint is valid, the tri-stimulus values are a convex combination of values on these lines, that is, \( X \) can be realized with the given illuminant if and only if it falls within the convex hull of these lines. Such a convex hull is a cone with vertex in the origin, called the color cone generated by the illuminant \( I \). Any value of the tri-stimulus that falls outside of this cone corresponds to a color that can't be generated under the given illuminant. In these circumstances, of course, the method for finding a feasible \( r \) would fail to converge. In practice, the tri-stimulus values from which one starts are often derived from a color generated under a certain illuminant and, if the illuminant \( I \) that we hypothesize for the
reconstruction is not wildly different from that under which the color was generated in the first place, the desired tri-stimulus values will fall within the color cone. This, at least, is the hypothesis that we will make throughout this paper.

III. THE METHOD

Suppose that we have solved the system (6) obtaining a weight vector \( \mathbf{w} \). From the discretization of (5)

\[
p^i = \sum_{h=1}^{b} \mathbf{w}_h \Psi_h^i
\]

we obtain the sampled reflectance \( r_0 = [p^1, \ldots, p^N] \). Define the clipping functions

\[
u_f(x) = \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x < 0 \end{cases}
\]

(10)

\[
u_+(x) = \begin{cases} x - 1 & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases}
\]

(11)

\[
u_-(x) = \begin{cases} -x & \text{if } x \leq 0 \\ 0 & \text{if } x > 0 \end{cases}
\]

(12)

then, applying the functions \( u \) to a vector componentwise, one can write

\[
r_0 = u_f(r_0) + u_+(r_0) = r_{0,f} + r_{0,+} + r_{0,-}
\]

(13)

Correspondingly, the tri-stimulus \( X \) can be decomposed from (4) as

\[
X = A_r = A_r = A_f + A_+ + A_- = X_{0,f} + X_{0,+} + X_{0,-}
\]

(14)

Consider the component \( X_{0,+} \): we can consider it as the datum of a new spectral estimation problem for which, using (6), we find the weights: \( w_{0,+} = C \cdot X_{0,+} \)

(15)

and from these the reflectance:

\[
\tilde{r}_{0,+} = \Psi w_{0,+} = \Psi C \cdot X_{0,+} = \Psi C \cdot A_r_{0,+}
\]

(16)

Similarly, the negative component \( X_{0,-} \) is the datum of an estimation problem from which we derive the spectrum:

\[
\tilde{r}_{0,-} = \Psi C \cdot X_{0,-} = \Psi C \cdot A_r_{0,-}
\]

(17)

Define now the reflectance: \( r_1 = r_{0,f} + \tilde{r}_{0,+} - \tilde{r}_{0,-} \)

(18)

From the previous construction, the following property follows immediately:

**Proposition 3.1.** The reflectances \( r_0 \) and \( r_1 \) are metameric under the given illuminant

The procedure can be repeated by decomposing

\[
r_1 = u_f(r_1) + u_+(r_1) - u_-(r_1) = r_{1,f} + r_{1,+} - r_{1,-}
\]

(19)

computing \( \tilde{r}_{1,+} = \Psi C \cdot A_r_{1,+} \) \( \tilde{r}_{1,-} = \Psi C \cdot A_r_{1,-} \)

(20)

then \( r_2 = r_{1,f} + \tilde{r}_{1,+} - \tilde{r}_{1,-} \)

(21)

and so on. In this way we obtain a sequence of metamer spectral

\[
r_n = [r_{n,f}, \ldots, r_{n,+}, r_{n,-}] \]

(22)

leaving, in the limit, a spectrum \( r_\infty = r_{\infty,f} \) which, by construction, is feasible. The reason why (22) holds will be considered in the next section but, before doing so, we should like to give an intuitive sense of why things work. Consider, for the sake of simplicity, only the portion \( r_{n,+} \) of the spectrum, the reasoning for \( r_{n,-} \) being analogous. Take a spectrum \( r_0 \) with values greater than 1, as in Figure 1.a, and isolate the part \( r_{0,+} \), as in Figure 1.b.

![Figure 1](image1)

**Figure 1.** (a) Example of an unfeasible reflectance function (b) The portion that exceed the maximum admissible value.

The unknown function \( \tilde{r}_{0,+} \) must be a metamer of \( r_{0,+} \), that is, in the continuum, it must be

\[
\int d\lambda I(\lambda) x^k(\lambda) \tilde{r}_{0,+} = \int d\lambda I(\lambda) x^k(\lambda) r_{0,+}
\]

where \( \Phi \) is the support of \( r_{0,+} \) that is, in general, smaller than \( \Lambda \) since \( r_{0,+} \) represents only a peak in an otherwise feasible spectrum. On the other hand, \( \tilde{r}_{0,+} \) is a combination of smooth basis functions and has in general the whole spectrum as support. In other words, in \( \tilde{r}_{0,+} \) we distribute the same “area” over a much wider support, so that the values of \( \tilde{r}_{0,+} \) will in general be smaller than those of \( r_{0,+} \) (see Figure 2).

![Figure 2](image2)

**Figure 2.** The spectrum of Figure 1.b and a metamer with a larger support and a smaller maximum value.

This is only a qualitative explanation to give the reader access to the intuition that was behind the inception of the method. In the following, we will put these intuitions on a more formal footing and determine the conditions of convergence of the algorithm. The complete algorithm is given in Table 1.

**Table 1.** The algorithm for the recovery of the reflectance function given the tristimulus values and the corresponding illuminant.
A. Smoothing the spectrum

Reflectance spectra of natural surfaces are smooth functions of wavelength. It comes that the basis functions computed on measured spectra are smooth functions as well (this is in general true, and comes from the choice of the basis set). Therefore, a reflectance spectrum obtained as a solution of the linear system in (6) is a smooth function. As discussed in Section 2, many studies have quantified the smoothness of natural reflectance functions in terms of content in the frequency domain. Observing the measured spectra of real surfaces, it is evident that such functions are characterized by a smooth curvature, as well. Estimated or synthesized spectra must exhibit this property in order to be practically realizable. Physically, that is, it is difficult to reproduce spectra with very high curvature values: each point of high curvature of the reflectance curve as a line in a 2D plane where the wavelength abscissa corresponds to \( x \) and the reflectance value corresponds to \( y \). In order to set an appropriate scale, reflectance values must be multiplied for 300, being 700-400 the wavelength domain of the curve. In this way, the ranges of values in the \( x \) and \( y \) coordinates have the same spatial extension. The procedure to round off the spectrum modifies the procedure of Table 1. Inside the \( \text{while} \) cycle, after the calculation of \( r_f \), \( r_+ \) and \( r_- \), a \( \text{smooth()} \) procedure should be inserted. One possible definition of the procedure is listed in Table 2, however other different smoothness measures can be adopted [Li2001][vanTrigt1990a].

In Table 2, \( N \) is the number of wavelength samples, and \( \alpha_{\text{min}} \) is the minimum angle between two segments composing the reflectance curve that ensures a smooth curvature. If we consider the example of Figure 4 and impose an angle limit of \( \frac{3}{4} \pi \), we obtain the rounded spectrum of Figure 5.

Unfortunately the transition from positive values of reflectance to the values equal to zero does not appear to be “natural”. To overcome this problem, we propose a method to “smooth” the spectrum where it appears to have an unfeasible shape, in order to increase its naturalness and make it physically realizable.

We define \( \text{shape feasibility} \) as the property of a spectrum to present a smooth curvature. The shape feasibility is a further requirement we impose to the final solution in addition to the \( \text{value feasibility} \) discussed in Section 2. We consider the reflectance curve as a line in a 2D plane where the wavelength abscissa corresponds to \( x \) and the reflectance value corresponds to \( y \). In order to set an appropriate scale, reflectance values must be multiplied for 300, being 700-400 the wavelength domain of the curve. In this way, the ranges of values in the \( x \) and \( y \) coordinates have the same spatial extension. The procedure to round off the spectrum modifies the procedure of Table 1. Inside the \( \text{while} \) cycle, after the calculation of \( r_f \), \( r_+ \) and \( r_- \), a \( \text{smooth()} \) procedure should be inserted. One possible definition of the procedure is listed in Table 2, however other different smoothness measures can be adopted [Li2001][vanTrigt1990a].

In Table 2, \( N \) is the number of wavelength samples, and \( \alpha_{\text{min}} \) is the minimum angle between two segments composing the reflectance curve that ensures a smooth curvature. If we consider the example of Figure 4 and impose an angle limit of \( \frac{3}{4} \pi \), we obtain the rounded spectrum of Figure 5.

![Figure 4](image-url)
TABLE 2. THE ALGORITHM TO INCREASE THE NATURALNESS OF A REFLECTANCE SPECTRUM.

\[
\text{smooth}(r_f, r_v, r_w) = \text{while}(i < N) do \\
\alpha_i = \tan^{-1} \left( \frac{d r_f}{d \lambda} \right) - \tan^{-1} \left( \frac{d r_f}{d \lambda} \right) + \Pi; \\
\text{if} (\alpha_i < \alpha_{\min}) \\
\quad r_i = \frac{r_i^{1} + r_i^{2}}{2}; \\
\quad \delta = r_i^{1} - r_i^{2}; \\
\text{if} (\delta < 0) r_i^{1} = \delta; \\
\text{if} (\delta > 0) r_i^{1} = \delta; \\
\text{return} r_f, r_v, r_w.
\]

Figure 5. The spectrum of Figure 4.b after the application of the procedure described in Table 2.

IV. CONVERGENCE

In this section, we put the intuitive explanation of the previous section on a more formal basis, and study the convergence of the method. We will start by introducing a class of linear transformations and their properties. Consider a linear transformation $\mathcal{R}^N \rightarrow \mathcal{R}^K$ ($K = 3$ being the case of interest here):

\[
X^k = \sum_{i=1}^{N} a_i^k r_i
\]

and let $r_i$ be expressed as the weighted sum of b vectors in $\mathcal{R}^N$, as in (9). Then, putting together (9) and (26), one obtains:

\[
X^k = \sum_{i=1}^{N} \sum_{k=1}^{b} a_i^k \psi_h^i w_h = \sum_{k=1}^{b} t_h^k w_h
\]

where $t_h^k = \sum_{i=1}^{N} a_i^k \psi_h^i$.

Let $\Lambda = [1, \ldots, N]$, then one can define the matrix $[\Lambda] = \begin{bmatrix} \Lambda^k \end{bmatrix}$ with $[\Lambda] : \mathcal{R}^b \rightarrow \mathcal{R}^K$. One can consider similar transformations in which the summation in (28) is extended only to a subset of the indices $\Lambda$. For instance, if $\Phi = [u, \ldots, v] , u \geq 1, v \leq N$, one has: $[\Phi] = \begin{bmatrix} \phi_h^k \end{bmatrix}$ and

\[
\phi_h^k = \sum_{i \in \Phi} a_i^k \psi_h^i
\]

The definition can be extended to any subset $\Phi \subseteq \Lambda$. Note that, for all the problems that we are considering here, $X^k \geq 0$, therefore:

\[
\|X\| = \|\Phi\| = \sum_{k=1}^{b} \phi_h^k w_h = \sum_{k=1}^{b} \phi_h^k w_h
\]

The following properties derive immediately from $a_i^k \geq 0$ and $\phi_h^k \geq 0$.

Lemma 4.1. For $\Phi, \Theta \subseteq \Lambda$,

\[
\Phi \cap \Theta = \emptyset \Rightarrow [\Phi \cup \Theta] = [\Phi] + [\Theta]
\]

In particular, for $a_i^k \geq 0$ $[\Lambda] = [\Phi] + [\overline{\Phi}]$ (32)

where $\overline{\Phi}$ is the complement of $\Phi$ in $\Lambda$.

Lemma 4.2. The 1-norm of $\sum_{i} \phi_h^k r^i$ is a norm for $r^i$, indicated as $\|r\|_1$ and, if $r^i = \sum_{k} w^k \phi_h^k$, and $[\Lambda]w = X$, then

\[
\|r\|_1 = \|\Lambda\|_1 w = \|X\|_1
\]

(33)

We can now consider the normalization procedure. Let $X_0$ be the given color (that is, the initial tri-stimulus of the procedure), and consider the tri-stimulus equations (4). Writing $r$ as in (9), the equation becomes $[A]w = X_0$. If $b = 3$, these equations can be solved for $w$,

\[
giving w = [\Lambda]^{-1} X_0 and the reflectance \]

\[
r_0 = [\Psi]^{-1} X_0
\]

with $\Psi = \{\phi_h^i\}$. The reflectance, in general, will not have a range included in $[0, 1]$. Here we assume, for the sake of simplicity, that $r^i \geq 0$, that is, that the only correction that we have to make are for those values $r^i > 1$. If there are parts with $r^i < 0$, one applies the transformation $r \leftarrow 1 - r$ to these parts, transforming them into parts with values grater than 1.

Define the set $\Phi_0 = \{r^i > 1\}$ and the functions $u^i$ and $v^i$, as in (10) and (11), respectively, so that

\[
r_0 = u_0(r_0) + u_+ (r_0) = r_{0,f} + r_{0,+}.
\]

Note that $r_{0,+} \geq 0$ and that the range of $r_{0,f}$ is in $[0, 1]$. With these definitions we have

\[
X_0^k = \sum_{i \in \Lambda} a_i^k r_0^i = \sum_{i \in \Lambda} a_i^k r_{0,f}^i + \sum_{i \in \Lambda} a_i^k r_{0,+}^i
\]

\[
\sum_{i \in \Phi_0} a_i^k r_{0,f}^i + \sum_{i \in \Phi_0} a_i^k (r_{0,+}^i - 1) \leq X_0^k + X_0^k
\]

(35)

Where the symbol $\leq$ means “equal by definition”. Note that $r_{0,+} > 1$ in $\Phi_0$, and that the reflectances are limited ($r_i < M$) so that one can write...
\[0 \leq \alpha^i = \frac{r^i_0 - 1}{r^i_0} < \frac{M - 1}{M} < 1\] (36)

In other words, one has
\[X_{0,+} = \sum_{i \in \Theta_0} a^i_0 r^i_0 = \sum_{i \in \Theta_0} \alpha^i_0 a^i_0 r^i_0 = \sum_{i \in \Theta_0} \alpha^i_0 r^i_0\] (37)

with \(\alpha^i_0 < 1\). The method now finds a reflectance function that generates \(X_{0,+}\). Expressing this reflectance as
\[\tilde{r}^i_0 = \sum_{k=1} w^h_k \psi^i_h\] (38)

one can find the coefficients \(w^h_k\) by solving the system \(X_{0,+} = [\Lambda]\tilde{w}^0_0\). On the other hand, from (37) and (9), one has \(X_{0,+} = \alpha_0 [\Phi_0] w_0\), leading to the relation between \(w_0\) and \(\tilde{w}^0_0\: [\Lambda]\tilde{w}^0_0 = \alpha_0 [\Phi_0] w_0\) (39)

In order to analyze this equation, we use the following property.

**Lemma 4.3.** Let \(u^i = \sum_{h} y^h \psi^i_h\) and \(v^j = \sum_{h} y^h \psi^j_h\) be two reflectance functions giving positive tri-stimulus values on the illuminant \(a^i_h\), and let \(x\) and \(y\) satisfy \([\Lambda]x = [\Phi]y\) with \(\Phi \subseteq \Lambda\). Then \(\|u^i\|_a \leq \|v^j\|_a\).

**Proof.** Since the functions are reconstructed from tri-stimulus values, which are positive by construction, eq. (32) applies. Consider the \(k\)th term of the equality:
\[([\Lambda] x)^k = ([\Phi] y)^k \leq ([\Phi] y)^k + ([\Phi] y)^k = ([\Lambda] y)^k\] (40)

where the last equality derives from lemma 4.1. So,
\[\|([\Lambda] x)^k\|_a = \sum_{k} \|([\Lambda] y)^k\|_a \leq \sum_{k} \|([\Lambda] y)^k\|_a = \|([\Lambda] y)^k\|_a\] (41)

From this and lemma 4.3 the result follows.

We now replace \(r_0 = r_{0,f} + r_{0,+}\) with \(r_1 = r_{0,f} + \tilde{r}_0\) and repeat the iteration. At the end we have
\[n_{i,f} = u_{i,f} (r_i) \] and \(n_{i,+} = u_{i,+} (r_i)\).
\[r_{i,f} \land r_{i,+} = u_{i,j} (r_i)\]

**Lemma 4.4.** In (42) it is \[\|\tilde{r}_i\|_a \leq \|r_0\|_a\]

**Proof.** All the values here are positive, so
\[\|\tilde{r}_i\|_a = \sum_{k=1} \|a^i_k \psi^i_k\|_a\] (43)

From this it follows that the proposition is true whenever \(r_{i,+}^i \leq \tilde{r}_i^i\) for all \(i\). We have the following cases:

i) \(r_{i,f}^i < 1\) \(\Rightarrow r_{i,+}^i = 0\) \(\Rightarrow r_{i,+}^i \leq \tilde{r}_i^i\)

ii) \(r_{i,f}^i = 1\) \(\land r_{i,+}^i = 1\) \(\Rightarrow r_{i,+}^i = \tilde{r}_i^i\)

iii) \(r_{i,f}^i = 1\) \(\land r_{i,+}^i = 1\) \(\Rightarrow r_{i,+}^i = r_{i,f}^i - r_{i,+}^i = r_{i,f}^i + \tilde{r}_i^i - r_{i,f}^i < \tilde{r}_i^i\)

where the symbol \(\land\) means logical conjunction.

**Lemma 4.5.** In the iteration, it is \[\|r_{i,+}\|_a \leq \|r_0\|_a\]

**Proof.** From lemma 4.3 and (39), it follows that \(\|r_0\|_a \leq \|r_0\|_a\). From this and lemma 4.4 the result follows. The main result of this section is given by the following theorem:

**Theorem 4.1.** Assume that a feasible spectrum exists, then the iteration
\[r(k) = u_f (r(k)) + u_+ (r(k)) \] with \(r(k) = \sum_i \psi^i_h r^i_h\) and \(\psi^i_h = \psi^i_h + \psi^i_h\) converges to a feasible solution.

**Proof.** Applying the previous lemmas to the \(k\)th iteration, one obtains
\[\|r(k)\|_a \leq \alpha \|r(k-1)\|_a\] (44)

and, iterating,
\[\|r(k)\|_a \leq \alpha^k \|r_0\|_a\] (45)

therefore
\[\lim_{k \to \infty} \|r(k)\|_a = 0\] (46)

so that the equilibrium point of the method is:
\[r_e = u_f (r_e)\] (47)

which, by definition of \(u_e\), is feasible. The requirement that the bases be positive is too strict for many a practical situation: one commonly used basis for spectral reconstruction, for instance, is obtained by principal components analysis on a suitable training set of reflectances. In this case, of course, there is no guarantee that the bases will be positive. It is possible, however, to prove that the same convergence result holds for a wider class of basis functions.

**Definition 4.1.** A basis \(\psi^i_h\) is said to be positive enough with respect to an illuminant \(a^i_h\) if
\[\sum_{h} \psi^i_h > 0\] (48)

for all \(h\). The most important property of positive enough functions, and the basis for the convergence result, is given by the following theorem:

**Theorem 4.2.** Let \(\psi^i_h\) be positive enough with respect to \(a^i_h\). Then it is possible to write \(\psi^i_h = \psi^i_h + \psi^i_h\), with \(\psi^i_h \geq 0\), and \(\sum a^i_h \psi^i_h = 0\) for all \(h\). Before proving this theorem, we introduce a lemma that will be used in the proof:

**Lemma 4.6.** Let \(A = \frac{1}{a^i_k} \frac{1}{a^j_k} a^i_k \frac{1}{a^j_k} \ldots a^i_k \frac{1}{a^j_k} \) be a matrix with \(N > K\) and \(rank(A) = K\). Then there is a matrix \(Z = \frac{1}{a^i_k} \frac{1}{a^j_k} \ldots \frac{1}{a^i_k} \) such that \(Z \geq 0\) and \(T = \frac{1}{a^i_k} \frac{1}{a^j_k} \ldots \frac{1}{a^i_k} \) is non-singular, where \(T = \sum_{i} a^i_k \frac{1}{a^i_k} \frac{1}{a^j_k} \ldots \frac{1}{a^i_k} \)

(49)

**Proof.** Since the matrix \(A\) has full rank, there are one or more sets of columns \(\mu = [i_1, \ldots, i_k]\)
are independent, since all the values $i_j$ for which $\mu_j \neq \mu_i$ are different, and $A \mathbf{Z}[\mu] = A \mu$. For $\alpha_1, \ldots, \alpha_M$ with $\alpha_\mu \geq 0$, define:

$$Z = \sum_\mu \alpha_\mu \mathbf{Z}[\mu]$$

so that: $A \cdot Z = \sum_\mu \alpha_\mu A \mu$.

any combination of $\alpha_\mu$ for which $\sum_\mu \alpha_\mu A \mu \neq 0$ will generate the required non-singular matrix $T = AZ$.

Proof of the theorem. Let $q_h^k = \sum_i a_i^k \psi_h^i$, and consider a basis of $K$ ($K = 3$ in our case) sampled functions $\xi_i^j$ such that the matrix $Z = \{\xi_i^j\}$ satisfies the previous lemma, and define:

$$\hat{\psi}_h^i = \sum_T \alpha_h^j \hat{\xi}_i^j$$

where the coefficients $\alpha_h^j$ are determined by the equalities:

$$q_h^k = \sum_i a_i^k \hat{\psi}_h^i = \sum_i \alpha_h^j \sum_i a_i^k \xi_i^j = \sum_i \alpha_h^j \xi_i^j$$

Because of lemma 4.6, it is possible to find positive functions $\hat{\xi}_i^j$ such that the matrix $T$ in the last equality is non-singular, and such that the $\alpha_h^j$ are positive. Defining $\hat{\psi}_h = \psi_h - \hat{\psi}_h$ completes the proof. Consider now a system such as (27):

$$X^k = \sum_i a_i^k r_i = \sum_i \left( \sum_h a_h^k \psi_h^i \right) w^h$$

Since $\psi_h^i = \psi_h^i + \hat{\psi}_h^i$, with $\sum_i a_i^k \hat{\psi}_h^i = 0$, the system can be rewritten as:

$$X^k = \sum_h \left( \sum_i a_i^k \hat{\psi}_h^i \right) w^h$$

In other words, the solution of the system depends only on the positive component of the basis $\psi_h^i$: the solution that we find is the same that we would find if, instead of the basis $\psi_h^i$, we were using the positive basis $\hat{\psi}_h^i$. The considerations that were made based on the positivity of $\psi_h^i$, therefore, still apply; in particular, eq. (39) still applies. From this point on, in the original demonstration, the result depend only on the positivity of expressions of the type $\sum_i a_i^k r_i$ and $\sum_i a_i^k \psi_h^i$, which hold by construction and by the assumption that the $\psi_h^i$ are positive enough. From this, therefore, we have the following result:

**Theorem 4.3.** The result of theorem 4.1 still holds if the basis $\psi_h^i$ is positive enough with respect to the illuminant $a_i^k$.

V. EVALUATION OF THE METHOD AND DISCUSSION

Our method is an iterative procedure that reduces an error at every step and, from a purely mathematical point of view, two questions arise quite naturally: how fast does it converge for a desired final error, and what is the final error if a certain execution time is allocated. In this section we report on our measurements of speed and quality comparing the results of our methods with those of another method that can be used to solve the same problem. Finally, we will try to determine the properties of the reconstructed spectra by looking at the color that the spectra determines under a change of illuminant. Before we explain our tests, we give a brief introduction to the use of linear programming to solve the spectral reconstruction problem, as linear programming is the method against which we measure our own.

A. Feasible spectra by linear programming

Linear programming is a well known optimization technique that can be used to solve constrained optimization problems where both the cost functional and the constraints are linear, and the solutions are required to be positive [Press1986]. That is, problems in which one seeks a vector $x \in \mathbb{R}^n$ such that

$$f = c^T x \quad c \in \mathbb{R}^n \text{ is minimal}$$

$$Ax = b \quad A \in \mathbb{R}^{m \times n} \quad b \in \mathbb{R}^m, \quad x_j \geq 0$$

where $A$, $b$, and $c$ are given constants. Note that this is not exactly the form of our reconstruction problem, since we don't quite have a cost to minimize and, in the case of strong feasibility, we have the additional constraint $x_i \leq 1$. Both these discrepancies can be eliminated by adopting the standard linear programming strategy of introducing additional variables. We can enforce the constraints $x_i \leq 1$ by introducing the variables $y_i$, $i = 1, \ldots, n$ and the constraints:

$$x_i + y_i = 1, y_i \geq 0$$

The introduction of these variables is equivalent to solving the linear programming problem with conditions

$$A'w = b', \quad w_i \geq 0, \quad w = \begin{bmatrix} x^T & y^T \end{bmatrix} \in \mathbb{R}^{2n}$$

$$A' = \begin{bmatrix} A & 0 \\ I_n & I_n \end{bmatrix} \in \mathbb{R}^{(n+n) \times 2n}$$

where $I_n$ is the $n \times n$ identity matrix, and $b' = \begin{bmatrix} b^T \vdots 1^T \end{bmatrix} \in \mathbb{R}^{m+n}$.
The lack of a function to be minimized can be solved by introducing the \( m + n \) additional variables \( z_i \), the conditions \( z = b' - A w \), i.e. \( z + A' w = b' \), and the cost \( f = \sum_i z_i = (1\ldots 1) \cdot z \). The cost is minimal for \( z = 0 \), and reaching this minimum entails finding a \( w \) for which the original condition is satisfied. In conclusion, in order to find a feasible solution, we solve the linear programming problem

\[
\begin{align*}
\min \ c^T v \\
A v &= b' \\
v_i &\geq 0
\end{align*}
\]

where \( v = \left[ x^T y^T z^T \right]^T \in \Re^{3n+m} \), \( b' = \left[ b^T | 1 \ldots 1 \right]^T \in \Re^{m+n} \), \( c = \left[ 2n \ldots n+m \right]^T \in \Re^{3n+m} \)

and \( A = \left[ I_n \ 0 \ -I_n \right]_{(m+n) \times (3n+m)} \) \quad (62)

This we solve using the standard techniques that can be found in the literature.

\section*{B. Feasible spectra by the method proposed}

In all, we do four separate measurements: execution time, colorimetric reconstruction, curvature of the reconstructed spectrum, and color evaluation under change of illuminant. The latter experiment follows a rather different experimental methodology than the first three, and will be considered separately in a different sub-section. While the ultimate quality gauge for an algorithm such as this one lies in its reconstruction quality, execution time is important for applications such as computer graphics, in which the spectral reconstruction will have to be repeated tens of thousands times per image. In order to give a general idea of what could be the advantages of the method in the area, we have include the execution time test. The two algorithms are implemented in C; for the simplex method we used the program provided in a reference book [Antia2001]—we also tried the implementation in a second reference book [Press1986] with virtually indistinguishable results. All reconstructions assumed the CIE D65 illuminant. The basis set adopted was obtained applying PCA to the Vrhe1’s dataset of objects reflectance spectra [Vrhe1994]. For each of the first three experiments we considered four different vector sizes: we started with the color matching functions and the illuminant covering the range from \( \lambda = 400\text{nm} \) to \( \lambda = 700\text{nm} \) in steps of 10nm, resulting in vectors of size \( N = 31 \), and then the experiment was repeated by down-sampling these vectors with steps 2, 3, and 4, obtaining vectors of dimension \( N = 16, 11, 8 \), respectively. Each experiment consisted in 30 trials, from which our statistics are derived; in each trial, we generated a random RGB color, converted it to XYZ, and run the trial on the two methods using the same color. For the \textit{time execution experiment}, the two algorithms were compiled using the lcc compiler with all optimizations turned off. In order to avoid interference from other processes running on the same computer, the time execution experiment was done under the DOS operating system. The two algorithms were compiled with the optimizer turned off to avoid having some particularity of the compiler used here affect the reproducibility of the results. Moreover, due to the unreliability of measuring short time intervals with the standard C timing functions, each trial consisted in the separate repetition of each algorithm \( 10^6 \) times. We indicate in the following the execution time as \( t_c \).

The colorimetric reconstruction experiment attempts to give a measure of the residual error in the reconstruction of the spectrum by computing back the tri-stimulus from which the reconstruction had started and comparing the reconstructed color with the original one. Given a color \( X \), a method is used to compute a feasible reflectance and then to this reflectance the tri-stimulus color. The outcome of the experiment is a reconstruction error:

\[
\Delta_X = \sum_{k} \left( X^k - \tilde{X}^k \right)^2 \right)^{1/2} \quad (64)
\]

The last experiment attempts to give a measure of the quality of the reflectance spectra obtained using the two methods. As discussed in Section 3, reflectance functions of natural surfaces or the spectra that we can physically construct are always smooth and relatively slowly varying. So, we take as a measure of the quality of the reconstructed spectra the average of the estimated curvature, that is:

\[
\Delta_c = \frac{1}{N-2} \sum_{i=2}^{N-1} \left| r^{i+1} - 2r^i + r^{i-1} \right| \quad (65)
\]

In this case, we didn’t apply the method in Section 3 to ensure the naturalness of reflectance functions.

\section*{C. Results and Discussion}

The results of our three tests are reported in tables 3 through 5. In all the tables, the first column contains the size of the vector used for that task, the second column the number of degrees of freedom for the analysis of the variance (which is equal to the number of trials minus one), the remaining eight columns contain the summary data for the two methods, and the last column contains the value of the t-test for that experiment. Table 3 contains the results of the time execution test.

\begin{table}[h]
\begin{center}
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
\textbf{N} & \textbf{dof} & \textbf{mean} & \textbf{std} & \textbf{min} & \textbf{max} & \textbf{t} & \textbf{p} \\
\hline
31 & 29 & 31.7 & 0.3 & 30 & 10 & 0 & 110 & 3.18 \\
31 & 29 & 31.0 & 0.4 & 29 & 10 & 0 & 110 & 2.96 \\
31 & 29 & 31.3 & 0.5 & 30 & 10 & 0 & 110 & 2.73 \\
31 & 29 & 31.8 & 0.6 & 30 & 10 & 0 & 110 & 2.50 \\
31 & 29 & 32.0 & 0.7 & 30 & 10 & 0 & 110 & 2.27 \\
31 & 29 & 32.3 & 0.8 & 30 & 10 & 0 & 110 & 2.04 \\
31 & 29 & 32.6 & 0.9 & 30 & 10 & 0 & 110 & 1.81 \\
31 & 29 & 32.9 & 1.0 & 30 & 10 & 0 & 110 & 1.58 \\
\hline
\end{tabular}
\end{center}
\caption{Results of the time execution test. Times are in ms. \( N \) is the dimensionality of the spectra being considered. Each trial consists of 30 randomly generated RGB colors, from which spectra are synthesized.}
\end{table}

The two tests with the larger vector sizes show statistically significant differences (\( t = 3.18 \) and \( t = 9.23 \), respectively), while for the shorter vector sizes the two methods perform...
equally. In the case of the larger vector size, the method proposed here appears to perform 2 to 4 times faster than linear programming, but this qualitative assessment hasn't been verified statistically. The large variance in the case N=31 is due to the fact that for a few colors both methods (the simplex and our own) fail to converge to the error specified but, when they converge, our method converges in a few iterations. Table 4 contains the results of the reconstruction error test. Here, for the two shortest vector sizes, the simplex method obtained always an error below the measurement threshold of $10^{-4}$. The difference between the reconstruction errors is not statistically significant although, by the average alone, in the case N=31 the method presented here appears to perform better.

Table 4: Results of the reconstruction error, that is the Euclidean distance between the actual XYZ values and the XYZ values of the reconstructed spectra (Eq. 64). All results are $10^3$.

<table>
<thead>
<tr>
<th>N</th>
<th>dof</th>
<th>µ</th>
<th>σ</th>
<th>min</th>
<th>max</th>
<th>µ</th>
<th>σ</th>
<th>min</th>
<th>max</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>29</td>
<td>1.3</td>
<td>4.1</td>
<td>0</td>
<td>21.4</td>
<td>64.8</td>
<td>258</td>
<td>0</td>
<td>1267</td>
<td>1.35</td>
</tr>
<tr>
<td>16</td>
<td>29</td>
<td>4.8</td>
<td>19.4</td>
<td>0</td>
<td>106</td>
<td>4.3</td>
<td>23.7</td>
<td>0</td>
<td>130</td>
<td>0.09</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>0.9</td>
<td>3.8</td>
<td>0</td>
<td>20.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.27</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>0.8</td>
<td>2.5</td>
<td>0</td>
<td>10.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.83</td>
</tr>
</tbody>
</table>

The reason for this lack of significance is in the high variance of the results, shown also by the high values of the maxima of the errors, due to the fact that occasionally the methods failed to converge in the allotted number of iterations. Note, however, that the maxima for the method presented here are lower than for linear programming, at least for N=16 and N=31. It appears, in other words, that the method presented here creates relatively rapidly an acceptable solution to the reconstruction problem so that, even if convergence is not reached within the allotted number of iterations, a reasonable solution is still found. Note that some of these problems might be solved by allowing a greater number of iterations but, as a matter of methodology, we decided to use here the same number of iterations that we used for the execution time test.

Table 5: Results of the curvature test.

<table>
<thead>
<tr>
<th>N</th>
<th>dof</th>
<th>µ</th>
<th>σ</th>
<th>min</th>
<th>max</th>
<th>µ</th>
<th>σ</th>
<th>min</th>
<th>max</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>29</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.05</td>
<td>8.2</td>
<td>3.96</td>
<td>0.8</td>
<td>14.9</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>29</td>
<td>0.14</td>
<td>0.08</td>
<td>0.02</td>
<td>0.34</td>
<td>8.0</td>
<td>8.48</td>
<td>3.2</td>
<td>38.2</td>
<td>5.1</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>0.4</td>
<td>0.24</td>
<td>0.06</td>
<td>1.01</td>
<td>8.2</td>
<td>9.20</td>
<td>0.9</td>
<td>46.2</td>
<td>4.7</td>
</tr>
<tr>
<td>8</td>
<td>29</td>
<td>0.55</td>
<td>0.26</td>
<td>0.13</td>
<td>1.04</td>
<td>22</td>
<td>15.6</td>
<td>4.2</td>
<td>59.4</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Finally, the curvature comparison is reported in Table 5. In this case, all the differences are statistically significant, and the method presented here produced results with a lower curvature, that is, smoother results, than linear programming. This was, of course, expected, since linear programming simply tries to find a vector that solves the optimization problem, without considering the vector as a sampled function and, therefore, without trying to enforce any relation between contiguous elements. The method presented here, on the other hand, derives the correction vector by sampling smooth basis functions and, therefore, produces smooth results. In order to give a visual explanation of the advantages of the method we propose with respect to a simple clipping of the initial solution obtained by a linear method, we report the plot of an experiment performed on a subset of the Munsell samples derived by [Wyszecky1982]. Figure 6 shows the section of the Munsell set of samples at the same level of brightness (value=5) in the CIE $a^*,b^*$ diagram. These data form the input for a recovery procedure based on the solution of equation (6) using the basis set published in [Fairman2004]. The recovered spectra, when reflectance values were out of the range [0,1] were clipped to ensure physical feasibility. The clipped spectra were then rendered under the CIE standard illuminant C, and the CIELAB coordinates of the colors computed. Results on the color opponent plane are reported in Figure 7a. Note that the gamut of colors has been significantly compressed. We then performed the same experiment but applying our method to correct the spectra instead of the clipping. Results are reported in Figure 7b. Note the difference in the gamut extension respect with the case of Figure 7a.

Figure 6. Subset (value=5) of the Munsell set of colors in the CIEL*a*b* color space. These data, converted in tristimulus values assuming the CIE Standard Illuminant C are the input to the reconstruction experiment.
The rationale for this procedure is that changes in illuminant should not change too much when the illuminant is changed. We expect that the reconstruction generated "reasonable" spectra, we should plot the result on a constant lightness plane. If the spectral and compute corresponding CIE(L*a*b*) values. We then under a different illuminant (the CIE standard illuminant A), standard illuminant C in our case). Then, we render the colors the spectral reconstruction using a certain illuminant (the CIE this section. The idea is the following: we consider a plane of colorimetric assessment, which is what we propose to do in previous sections tried to measure the extent to which this "physical" characteristic of naturally occurring spectra was present in the spectra reconstructed with the different methods. In order to have a better idea of the behavior of the spectra, however, it is desirable to have some form of colorimetric assessment, which is what we propose to do in this section. The idea is the following: we consider a plane of the Munsell color system with constant value (we have taken the plane with value equal to 5) thus selecting a subset of samples in the Munsell dataset. For each selected color, we do the spectral reconstruction using a certain illuminant (the CIE standard illuminant C in our case). Then, we render the colors under a different illuminant (the CIE standard illuminant A), and compute corresponding CIE(L*a*b*) values. We then plot the result on a constant lightness plane. If the spectral reconstruction generated "reasonable" spectra, we should expect that the pattern that we observe in the Munsell plane should not change too much when the illuminant is changed. The rationale for this procedure is that changes in illuminant

In the first three experiments, for the sake of control and repeatability, we used a random set of colors and a fixed illuminant. In this situation, there is no guarantee that any natural object, or "natural" reflectance will be able to generate a given color under the assumed illuminant. There is some indications that the areas in which the method shows a slow convergence correspond to rather irregular reflectance functions: since the reflectance functions that the method builds are created by repeated addition and clipping of linear combinations of the bases, which are very regular and smooth, rendering a highly irregular reflectance is very difficult, and it may require many iterations. Our method, on the other hand, starts with a non-feasible solution and builds increasingly feasible approximations of it. If the method is interrupted before it converges, it will produce a feasible reflectance by "clipping" whatever approximation it has at that time. This behavior tends to provide reasonable (although not strictly metameric) approximations of the true reflectance and to provide acceptable solutions even in cases in which convergence is not reached. We owe to this, for instance, the fact that even the points in Figure 8 for which the iterations were interrupted are not placed in wildly wrong positions of the plane. This seems to be the case, on the other hand, with the simplex method: whenever the method can't converge in the allotted number of iterations, the approximation of the reflectance provided is strongly non-metameric to the initial one, and the corresponding point can be virtually everywhere in the color plane.

D. Change of illuminant

There is, of course, more to spectral feasibility than having the range included in [0,1]; the spectra of natural objects tend to be smooth, for example. The curvature experiment in the previous sections tried to measure the extent to which this "physical" characteristic of naturally occurring spectra was present in the spectra reconstructed with the different methods. In order to have a better idea of the behavior of the spectra, however, it is desirable to have some form of colorimetric assessment, which is what we propose to do in this section. The idea is the following: we consider a plane of the Munsell color system with constant value (we have taken the plane with value equal to 5) thus selecting a subset of samples in the Munsell dataset. For each selected color, we do the spectral reconstruction using a certain illuminant (the CIE standard illuminant C in our case). Then, we render the colors under a different illuminant (the CIE standard illuminant A), and compute corresponding CIE(L*a*b*) values. We then plot the result on a constant lightness plane. If the spectral reconstruction generated "reasonable" spectra, we should expect that the pattern that we observe in the Munsell plane should not change too much when the illuminant is changed. The rationale for this procedure is that changes in illuminant

The patterns show that the distortion caused by the simplex method are significantly larger than those caused by the present method. The main distortion of the present method, apart from translation, rotation, and shear (shear is in any case much less pronounced for our method than for the simplex) appears to be a compression of the two "long arms" of the Munsell plane corresponding to low values of b* and negative values of a*. In many cases, this is due to a slow convergence of the method: our test program would cut the method after 30 iterations, even if the point of convergence hadn't been reached. By comparison, the simplex method causes heavy distortion of the whole portion b*>0 of the Munsell plane, to the point that the pattern is almost impossible to detect in this region.

VI. Conclusion

In this paper we proposed a solution to correct the outcome of reflectance recovery methods, in order to ensure the physical feasibility and naturalness of estimated reflectance functions. Our method is iterative, and converges to a feasible metamer of the initial recovered reflectance function. As future research we would like to investigate the connection between...
the type of surfaces, the basis functions adopted, and our method.

REFERENCES