Constrained Nonlinear Optimization Approaches to Color-Signal Separation

Po-Rong Chang, *Member, IEEE,* and Tsung-Hsieh Hsieh

Abstract— **The process of separating a color signal into illumination and surface reflectance components is a fundamental issue in color reproduction and constancy. This color-signal separation can be carried out by minimizing the error in the least squares fit of the product of the illumination and the surface spectral reflectance to the actual color signal. Moreover, when taking in account the physical realizability constraints on both the surface reflectance and illumination, the feasible solutions to the nonlinear least-squares problem should satisfy a number of linear inequalities. Four distinct novel optimization algorithms are presented to employ these constraints to minimize the nonlinear least squares fitting error. The first approach, which is based on Ritter's superlinear convergent method, provides a computationally superior algorithm to find the minimum solution to the nonlinear least-squares error problem subject to linear inequality constraints. Unfortunately, this gradient-like algorithm may sometimes be trapped at a local minimum or become unstable when the parameters involved in the algorithm are not tuned properly. The remaining three methods are based on the stable and promising global minimizer called simulated annealing. The annealing algorithm can always find the global minimum solution with probability one, but its convergence is extremely slow. To tackle this difficulty, a cost-effective variable-separable formulation based on the concept of Golub and Pereyra is adopted to reduce the nonlinear least-squares problem to be a small-scale nonlinear least-squares problem whose solution state space is less than that of the original space. It will be shown that the computational burden is reduced by an order of magnitude. The computational efficiency can be further improved when the original Boltzman generating distribution of the classical annealing is replaced by the Cauchy distribution. Finally, a number of test samples are conducted to verify the effectiveness of the proposed methods.**

I. Introduction

THE spectral power distribution of the ambient light and the surface spectral reflectance of objects in a still image are the two primary physical factors that influence the color appearance of the objects in the image. The information obtained during image acquisition and display of a synthetic image is called the color signal. The term color signal will be used to describe the spectral power distribution of the light arriving at the recording device or the human eye. In a natural scene, the color signal is generally computed as the product of the spectral power distribution of the light incident on an object and the surface reflectance of that object.

Manuscript received July 10, 1992; revised October 11, 1993. This work was supported in part by the National Science Council, R.O.C., under Grant NSC-81-0404-E-009-027. The associate editor coordinating the review of this paper and approving it for publication was Dr. Frederick Mintzer.

The authors are with the Department of Communication Engineering, National Chiao-Tung University, Hsin-Chu, Taiwan, R.O.C.

IEEE Log Number 9406885.

In the analysis of color images, the perceived surface color descriptors in the acquired image are determined by the color signal. More precisely, the recovery of color descriptors is always recognized as extracting the underlying surface spectral reflectance and ambient light spectral power distribution from the color signal. The knowledge of surface spectral reflectance can improve color reproduction to render the scene's color best.

Wandell [1] and Maloney [2] proposed a method of recovering both surface reflectance and illumination that requires solving a set of equations based on the receptor values from several regions of different colors. Recently, Ho *et al.* [3] showed that the method has several primary limitations. One of them is that Wandell's system cannot capture the richness of surface color with the information of the lower dimensionality. To extract the illumination and surface reflectance with higher dimensionality, it requires a sampling of the entire (visible) color-signal spectrum. In [3] and [4], the spectral information of color signal can be derived from the chromatic aberration inherent in lens systems. Once the sampled color signal has been determined, the process of dividing it into one component due to the illumination and the second due to the reflectance is executed by minimizing a nonlinear least-square error between the sampled color signal and the product of both illumination and reflectance.

Ho *et al.* [3] showed that this nonlinear least-squares solution provides a good fit to illumination and surface reflectance individually. More precisely, their separation algorithm derives least squares approximations to the actual illumination and reflectance based on Judd's [5] and Cohen's [6] finite-dimensional linear models, respectively. The error bound of each approximation is dependent on the accuracy of both linear models. The correct color signal separation is achieved when both the finite-dimensional models describe illumination and reflectance exactly, but the primary limitation of their method is that the recovered components may violate the physical constraints. For example, the recovery procedure may give rise to estimates of the surface reflectance functions that are not physically realizable, that is, that contain negative values of surface reflectivity. In this paper, a constrained nonlinear least squares formulation is introduced by taking into account the physical constraints on both illumination and reflectance. Section II shows that these constraints can be mathematically expressed as a number of linear inequalities. A computationally superior algorithm based on the concept of Ritter's superlinearly convergent method [15] is applied to find the minimum solution subject to linear inequality constraints

and without computation of second-order derivatives. More details will be discussed in Section III. Unfortunately, this gradient-like algorithm sometimes may become trapped in a local minima. Moreover, it becomes unstable when the parameters involved in the algorithm are not tuned properly.

In constrast to the superlinearly convergent method, a method based on simulated annealing [20] can reduce the possibility of being trapped in a local minima. The annealing algorithm is a stochastic optimization algorithm derived from Monte Carlo methods in statistical mechanics. Corana *et al.* [26] showed that the convergence with probability one sense to a global minimum is provable. To implement the simulated annealing algorithm for the problem of separating the color signal successfully, it requires generating the interior points belonging to a convex polytope with a uniform distribution. This can be achieved by using the hit-and-run algorithm [12]. The algorithm chooses the current interior point on the line segment connecting the two hitpoints that are the intersections between the boundary of the convex polytope and a line formed by the previous interior point and a random direction vector. Moreover, it guarantees that the generated random sequence of interior points has the limiting uniform distribution.

Undoubtedly, the linearly constrained simulated annealing can be utilized to find the global minimum solution over the convex polytope whose dimensionality is proportional to the total number of weighting coefficients corresponding to both the linear models. It can be shown that the convergence of the annealing algorithm becomes extemely slow when the dimensionality grows larger. To improve the computational efficiency, a variable-separable formulation based on Golub and Pereyra [11] is adopted to reduce the nonlinear least-squares problem to a small-scale nonlinear least squares problem. This can divide the solution state space of the specific annealing algorithm into a small-scale state space. A new cost function known as the variable projection functional is introduced and defined in the small-scale solution state space. Here, a minimum point defined in this variable projection functional is obtained by the same annealing algorithm on its associated state space whose dimensionality is smaller. According to Theorem 2 in Section V, once the minimum solution to the variable projection functional has been obtained, the minimum point to the original problem is determined. It will be shown that the variable-separable formulation can reduce the computational burden of the original annealing algorithm by an order of magnitude. At the end of Section V, we will discuss the fast simulated annealing [29] whose generating probability distribution is a Cauchy distribution instead of the original Boltzman distribution. The computational efficiency is further improved. Finally, a number of illustrated examples, experimental results, and comparisons will be discussed in the last section.

II. A Constrained Nonlinear Optimization Formulation for Color Signal Separation

Ho *et al.* [3] showed that the problem of separating a color signal into illuminant and surface reflectance components can be identified as solving an unconstrained nonlinear least squares. Unfortunately, their method cannot guarantee that both the recovered components satisfy the physical constraints. The resultant illuminant and reflectance will become infeasible when they violate their corresponding physical constraints. To overcome this difficulty, our formulation includes the consideration of physical constraints. In a natural scene, the color signal is generally computed as the product of the spectral power distribution of the light incident on an object and the surface reflectance. Ho *et al.* [3] proposed a method to derive the spectral information of color signal from chromatic aberration. The color signal spectra can also be estimated, based on the response values of photoreceptors by a technique of Hermite interpolation polynominal [10] or principal component approximate basis [7], [28]. Once the color signal has been determined, Ho *et al.* [3], [4] indicated that the process of separating a color signal is achieved by performing a nonlinear least squares or solving a set of equations. Moreover, Wandell [1] gave a condition that ensures the uniqueness of the decomposition. As mentioned above, none of their methods consider the validity of the physical constraints on both illuminant and surface reflectance. For example, the recovered feasible surface reflectance should have their values fall between zero and one according to the conventional definition of reflectance [1]. The spectral power distribution of the illuminant is never less than zero nor greater than the upper limitation of the light source. A constrained nonlinear optimization formulation based on finite-dimensional linear models will be introduced to extract the feasible illuminant and reflectance from a given color signal below.

A. Finite-Dimensional Approximate Models for Illuminant and Surface Reflectance

In general, finite-dimensional models will not repesent the spectral power distribution of illuminant and the surface spectral reflectance exactly, although they should approximate them. Cohen [6] and Maloney [2] showed that the surface spectral reflectance curve of natural objects are usually reasonably smooth and continuous over the feasible spectrum (400–700 nm). Many experiments on empirical surface spectral reflectance show that most of them can be modeled by a finite-dimensional linear representation using only a few basis functions. For example, Cohen [6] found that over 99% of the variance of the spectral reflectance functions of the Munsell chips can be expressed using only three principal components. This analysis has been confirmed and extended by Maloney [2]. Higher dimensions result in better approximation, yet three functions still suffice when the filtering effect of cone response functions is taken into account. The estimate of the surface spectral reflectance corresponding to position $\boldsymbol{p}$ in sensor array is expressed as

$$S^{\boldsymbol{p}}(\lambda) \cong \sum_{j=1}^{n} \beta_j^{\boldsymbol{p}} s_j(\lambda) \tag{1}$$

where $s_j(\lambda)$ is the jth basis function, β_j is its associated position-dependent coefficient, and $n = 3$. For simplicity, we

neglect the position-dependent index $\boldsymbol{p}$ and obtain

$$S(\lambda) \cong \sum_{j=1}^{n} \beta_j s_j(\lambda). \tag{2}$$

Consider approximating the spectral power distribution of light. Judd *et al.* [5] reported that nearly all of the variations in the spectral power distribution of natural daylight can be described using a linear model consisting of three terms. This was confirmed in the later studies by Satri and Das as well as Maloney [2]. In addition, Maloney [2] has shown that the principal components that describe the observed variations in daylight also describe the variations across another class of light source: the blackbody radiators. In addition, it is assumed that the ambient light remains constant over the scene. This implies that the illuminant is independent of the location in an array. From these results, the spectral power distributions of lights can be characterized as

$$E(\lambda) \cong \sum_{i=1}^{m} \alpha_i e_i(\lambda) \tag{3}$$

where $e_i(\lambda)$ is the ith basis function, α_i is its associated coefficient, and $m = 3$.

Hence, the color signal $I(\lambda)$ at a position in a sensory array can be expressed as a linear combination of $m \times n$ functionals.

$$\begin{aligned} I(\lambda) = E(\lambda)S(\lambda) &\cong \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha_i\beta_j)(s_j(\lambda)e_i(\lambda)) \\ &\cong \sum_{i=1}^{m}\sum_{j=1}^{n}\rho_{ij}\phi_{ij}(\lambda) \end{aligned} \tag{4}$$

where $\phi_{ij}(\lambda) = e_i(\lambda)s_j(\lambda)$, and $\rho_{ij} = \alpha_i\beta_j$.

It should be noted that all the $m \times n(= 9)$ product-pair functionals $\phi_{ij}(\lambda)$ form a linearly independent set.

B. Nonlinear Least Squares Problems with Linear Inequality Constraints

Determining the illuminant and surface spectral reflectance can be described as the attempt to select their corresponding coefficients α_i and β_j of linear models that minimize the metric between the given actual color signal $I(\lambda)$ and the unknown model $\hat{I}(\lambda)$ formed by the product of linear models of both $E(\lambda)$ and $S(\lambda)$ over the visible spectrum.

$$\begin{aligned} d(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{k=1}^{N}(I(\lambda_k) - \hat{I}(\lambda_k))^2 \\ &= \|\vec{I} - \hat{I}\|^2 \end{aligned} \tag{5}$$

where $\hat{I}(\lambda_k) = \sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\beta_j e_i(\lambda_k)s_j(\lambda_k)$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T$, $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_n]^T$, $\vec{I} = [I(\lambda_1), I(\lambda_2), \ldots, I(\lambda_N)]^T$, λ_k is the kth sampling wavelength over the visible spectrum, N is the number of sampling, $\| \bullet \|$ represents the Eucledian norm, and T denotes the matrix transpose.

Ho *et al.* [3] showed that there are an infinite number of solutions to the minimization problem. Since the α_i and β_j unknowns appear in products, we can solve for them only up to a multiplicative scaling factor. To do so, the illuminant function $E(\lambda)$ is suggested to be normalized such that $\alpha_1 = 1$. Then, we have unique solution for all α_i and β_j. In other words, the unique derivation of the best estimated illuminant and reflectance is guaranteed except for a scaling factor applied to normalize α_1 to 1. Unfortunately, the desired feasible solution vector $[\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_n]^T = [\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T]^T$ cannot be obtained by minimizing the metric of (5) directly. Since the surface reflectance function $S(\lambda)$ have the property that their values fall between zero and one, these reflectance weighting coefficients β_j should satisfy the constraint of boundedness and belong to the feasible reflectance coefficient set $\Psi = \{[\beta_1, \beta_2, \ldots, \beta_n]^T \mid 0 \le \sum_{j=1}^{n}\beta_j s_j(\lambda_k) \le 1, 1 \le k \le N\}$. Similarly, the feasible region for the illumination is found to be $\Phi = \{[\alpha_1, \alpha_2, \ldots, \alpha_n]^T \mid I(\lambda_k) \le \sum_{i=1}^{m}\alpha_i e_i(\lambda_k) \le \zeta, 1 \le k \le N, \zeta =$ the upper bound of the illuminant$\}$ by taking into account the fact that $E(\lambda) = I(\lambda)/S(\lambda)$ and $0 \le S(\lambda) \le 1$.

In summary, the constrained nonlinear least squares can be stated as follows:

$$\begin{aligned} \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} d(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{k=1}^{N}(I(\lambda_k) - \sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\beta_j e_i(\lambda_k)s_j(\lambda_k))^2 \\ &= \left\| \vec{I} - \sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\beta_j\vec{es}_{ij} \right\|^2 \end{aligned} \tag{6}$$

where $\vec{es}_{ij} = [e_i(\lambda_1)s_j(\lambda_1), e_i(\lambda_2)s_j(\lambda_2), \ldots, e_i(\lambda_N)s_j(\lambda_N)]^T$.

Subject to a set of $4N$ inequality constraints and a equality constraint

$$\begin{aligned} &(1)\ \alpha_1 = 1 \\ &(2)\ I(\lambda_k) \le \sum_{i=1}^{m}\alpha_i e_i(\lambda_k) \le \zeta, 1 \le k \le N \end{aligned} \tag{7.a}$$

and

$$(3)\ 0 \le \sum_{j=1}^{n}\beta_j s_j(\lambda_k) \le 1, 1 \le k \le N \tag{7.b}$$

or equivalently, the sets of $(4N+2)$ inequality constraints are given by

$$\hat{\Phi} = \Phi|_{\alpha_1=1} = \{\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T \mid \Sigma\boldsymbol{\alpha} \le \vec{E}^+\} \tag{8}$$

and

$$\Psi = \{\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_n]^T \mid \Lambda\boldsymbol{\beta} \le \vec{S}^+\} \tag{9}$$

where $\vec{E}^+ = [1, -1, \zeta, \zeta, \ldots, \zeta, -I(\lambda_1), \ldots, -I(\lambda_N)]^T$, and $\vec{S}^+ = [1, 1, \ldots, 1, 0, 0, \ldots, 0]^T$ are the $(2N+2)\times 1$ and $2N \times 1$ matrices, respectively. $\vec{e}_k = [e_1(\lambda_k), e_2(\lambda_k), \ldots, e_m(\lambda_k)]^T$ and $\vec{s}_k = [s_1(\lambda_k), s_2(\lambda_k), \ldots, s_n(\lambda_k)]^T$ are the $m \times 1$ and

$n \times 1$ column vectors, respectively.

$$\Sigma = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 \\ e_1(\lambda_1) & e_2(\lambda_1) & \cdots & e_m(\lambda_1) \\ e_1(\lambda_2) & e_2(\lambda_2) & \cdots & e_m(\lambda_2) \\ \vdots & \ddots & \cdots & \\ e_1(\lambda_N) & e_2(\lambda_N) & \cdots & e_m(\lambda_N) \\ -e_1(\lambda_1) & -e_2(\lambda_1) & \cdots & -e_m(\lambda_1) \\ -e_1(\lambda_2) & -e_2(\lambda_2) & \cdots & -e_m(\lambda_2) \\ \vdots & \ddots & \cdots & \\ -e_1(\lambda_N) & -e_2(\lambda_N) & \cdots & -e_m(\lambda_N) \end{bmatrix}_{(2N+2)\times m}$$

$$= \begin{bmatrix} (\vec{\epsilon}_{m\times 1})_1^T \\ -(\vec{\epsilon}_{m\times 1})_1^T \\ \vec{e}_1^T \\ \vec{e}_2^T \\ \vdots \\ \vec{e}_N^T \\ -\vec{e}_1^T \\ -\vec{e}_2^T \\ \vdots \\ -\vec{e}_N^T \end{bmatrix}_{(2N+2)\times m} \tag{10}$$

and

$$\Lambda = \begin{bmatrix} s_1(\lambda_1) & s_2(\lambda_1) & \cdots & s_n(\lambda_1) \\ s_1(\lambda_2) & s_2(\lambda_2) & \cdots & s_n(\lambda_2) \\ \vdots & \ddots & \cdots & \\ s_1(\lambda_N) & s_2(\lambda_N) & \cdots & s_n(\lambda_N) \\ -s_1(\lambda_1) & -s_2(\lambda_1) & \cdots & -s_n(\lambda_1) \\ -s_1(\lambda_2) & -s_2(\lambda_2) & \cdots & -s_n(\lambda_2) \\ \vdots & \ddots & \cdots & \\ -s_1(\lambda_N) & -s_2(\lambda_N) & \cdots & -s_n(\lambda_N) \end{bmatrix}_{2N\times n}$$

$$= \begin{bmatrix} \vec{s}_1^T \\ \vec{s}_2^T \\ \vdots \\ \vec{s}_N^T \\ -\vec{s}_1^T \\ -\vec{s}_2^T \\ \vdots \\ -\vec{s}_N^T \end{bmatrix}_{2N\times n} \tag{11}$$

where $(\vec{\epsilon}_{m\times 1})_i$ is an $m \times 1$ vector with unity in its ith coordinate and zeros elsewhere.

Lemma 1: $\hat{\Phi}$ and Ψ are identified as the convex polytopes that can be expressed as the intersection of a finite number of closed half spaces. Clearly, Ψ contains a trivial feasible solution $\vec{0} = [0, 0, \ldots, 0]^T$. Moreover, the more compact representations for $\hat{\Phi}$ and Ψ with fewer inequalities are obtained by eliminating the redundant inequalities [17].

In many cases, the systems of inequalities (7.a) and (7.b) used to define $\hat{\Phi}$ and Ψ may not be the simplest, and it may be possible to find the other systems having fewer inequalities while defining the same feasible solution sets $\hat{\Phi}$ and Ψ. A system of linear inequalities can be simplified by eliminating the redundant inequalities. This leads to a new equivalent system having the same number of variables but fewer inequalities. These redundent inequalities that are deleted from a system without changing its set of feasible solutions can be expressed as a linear combination of the others. Methods to eliminate such constraints have been proposed by many researchers. Most of them are based on the simplex method [17]. For ease of notation, both Σ and Λ defined in $\hat{\Phi}$ and Ψ are still used to represent the nonredundant inequalities that are deduced from the original inequalities.

C. Accuracy of the Nonlinear Least-Squares Solutions

This section discusses the accuracy of the nonlinear least squares solutions to the color signal separation problem. Ho *et al.* [3] introduced the following theorem to describe how the coefficients α_i and β_j representing the best fits to a color signal $I(\lambda)$ and obtained by solving the constrained nonlinear least squares problem provide a good fit to $E(\lambda)$ and $S(\lambda)$ individually. In other words, the error bound of the fit to each of both $E(\lambda)$ and $S(\lambda)$ is proportional to the errors involved in fitting illuminant and surface reflectance separately with the finite-dimensional model. The details are discussed in the following theorem.

Theorem 1 [3]—The Error Bounds for Color Signal Separation: Let $E(\lambda)$, $S(\lambda)$, $e_i(\lambda)$, and $s_j(\lambda)$ be bounded functions, and let $I(\lambda)$ be a given normalized color signal. The color signal is normalized if the illuminant contained in the signal is normalized. The errors involved in fitting both $E(\lambda)$ and $S(\lambda)$ are given by $\epsilon_E = \min_{\boldsymbol{\alpha}\in\hat{\Phi}}\{\|\sum_{i=1}^{n}\alpha_i\vec{e}_i - \vec{E}\|\}$ and $\epsilon_S = \min_{\boldsymbol{\beta}\in\Psi}\{\|\sum_{j=1}^{m}\beta_j\vec{s}_j - \vec{S}\|\}$. Let x_i, $1 \le i \le m$ and $x_1 = 1$, and let y_j, $1 \le j \le n$ be those values of α_i and β_j so that a best fit is provided for the color signal:

$$\left\|\sum_{i=1}^{m}\sum_{j=1}^{n}x_iy_j\vec{es}_{ij} - \vec{I}\right\| \le \min_{\boldsymbol{\alpha}\in\hat{\Phi},\boldsymbol{\beta}\in\Psi}\left\|\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\beta_j\vec{es}_{ij} - \vec{I}\right\|. \tag{12}$$

If all the product-pair functions $e_i(\lambda)s_j(\lambda)$, $1 \le i \le m$, $1 \le j \le n$ are linearly independent, then there exist constants C and D such that

$$\left\|\sum_{i=1}^{m}x_i\vec{e}_i - \vec{E}\right\| \le C(\epsilon_E + \epsilon_S) \tag{13}$$

and

$$\left\|\sum_{j=1}^{n}y_j\vec{s}_j - \vec{S}\right\| \le D(\epsilon_E + \epsilon_S) \tag{14}$$

where $\vec{E} = [E(\lambda_1), E(\lambda_2), \ldots, E(\lambda_N)]^T$, $\vec{S} = [S(\lambda_1), S(\lambda_2), \ldots, S(\lambda_N)]^T$, $\vec{e}_i = [e_i(\lambda_1), e_i(\lambda_2), \ldots, e_i(\lambda_N)]^T$, $\vec{s}_j = [s_j(\lambda_1), s_j(\lambda_2), \ldots, s_j(\lambda_N)]^T$, $\vec{es}_{ij} = [e_i(\lambda_1)s_j(\lambda_1), \ldots, e_i(\lambda_N)s_j(\lambda_N)]^T$, and $\hat{\Phi} = \Phi\,|_{\alpha_1=1}$.

Theorem 1 implies that the modelling errors of the linear models to both $E(\lambda)$ and $S(\lambda)$ based on the coefficients x_i and y_i tend to zero when the spectral power distribution of the illuminant and the surface spectral reflectance are exactly represented by the finite-dimensional linear models. Otherwise, these modeling errors are bounded by the weighted sums of

the error bounds for linear model approximations to both E and S. Note that theorem 1 is valid if the illuminant involved in the given color signal $I(\lambda)$ is normalized. Throughout this paper, we assume that both $I(\lambda)$ and $E(\lambda)$ are normalized such that their α_1's are identical to unity. In the next section, we will provide an systematic approach to obtain these optimal coefficients x_i and y_j.

III. Superlinearly Convergent Minimization Formulation

Ritter [15] proposed an extension of the superlinearly convergent method for minimization problems with linear inequality constraints that is suitable for the implementation of the color signal separation optimization problem. It can be applied under the same general assumptions as any method of steepest descent, even if the superlinear convergence assumptions [15] are not satisfied. The general assumptions of steepest descent methods are recognized such that cost function $F(\boldsymbol{u})$ should be twice continuously differentiable over a convex region. Under these fairly general assumptions, the algorithm generates a sequence of points that converges near-superlinearly to a minimizer of $F(\boldsymbol{u})$ without computation of second-order derivatives. Certainly, the superlinear convergence is achieved when the stringent assumptions on the second-order derivatives [15] are satisfied.

To implement the color signal separation problem, the constrained nonlinear least-squares of (6) and its associated variables should be restated

$$\min_{\boldsymbol{u}} F(\boldsymbol{u}) \left(= \sum_{k=1}^{N} (I(\lambda_k) - \sum_{i=1}^{m} \sum_{j=m+1}^{m+n} u_i u_j e_i(\lambda_k) s_{j-m}(\lambda_k) \right)^2 = \|\vec{I} - \vec{\eta}(\boldsymbol{u})\|^2) \tag{15}$$

subject to

$$A\boldsymbol{u} \leq \boldsymbol{b} \tag{16}$$

where $\boldsymbol{u} = [\boldsymbol{u}_1^T, \boldsymbol{u}_2^T]^T = [\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T]^T$, $\boldsymbol{u}_1 = [u_1, u_2, \ldots, u_m]^T$, and $\boldsymbol{u}_2 = [u_{m+1}, u_{m+2}, \ldots, u_{m+n}]^T$

$$u_i = \begin{cases} \alpha_i & 1 \leq i \leq m \\ \beta_{i-m} & m+1 \leq i \leq m+n \end{cases} \tag{17}$$

and there are two different explicit triple matrix-vector product expressions for representing $\vec{\eta}(\boldsymbol{u})$, which are given by

$$\vec{\eta}(\boldsymbol{u}) = E(\boldsymbol{u}_1) S^0 \boldsymbol{u}_2 \tag{18.a}$$

and

$$= S(\boldsymbol{u}_2) E^0 \boldsymbol{u}_1 \tag{18.b}$$

where $E(\boldsymbol{u}_1) = \text{diag}(\sum_{i=1}^{m} u_i e_i(\lambda_1), \sum_{i=1}^{m} u_i e_i(\lambda_2), \ldots, \sum_{i=1}^{m} u_i e_i(\lambda_N))$, $S(\boldsymbol{u}_2) = \text{diag}(\sum_{i=m+1}^{m+n} u_i s_{i-m}(\lambda_1), \sum_{i=m+1}^{m+n} u_i s_{i-m}(\lambda_2), \ldots, \sum_{i=m+1}^{m+n} u_i s_{i-m}(\lambda_N))$, $E^0_{N \times m} = [\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_N]^T$, $S^0_{N \times n} = [\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_N]^T$ and $\text{diag}(d_1, d_2, \ldots, d_N)$ is meant to be the $N \times N$ diagonal matrix whose (i,i)-entry equals d_i for $1 \leq i \leq N$.

By comparing (18.a) and (18.b), we observe that there is a definite symmetry (i.e., that these equations are similar but not quite identical in form). In fact, this symmetry leads to a property of duality between these two expressions.

The strategy of the superlinearly convergent method is based on the following gradient-like iterative algorithm of the form [17]

$$\boldsymbol{u}_{j+1} = \boldsymbol{u}_j - \sigma_j \boldsymbol{v}_j \tag{19}$$

where σ_j and $\boldsymbol{v}_j$ are the adjustable scalar step size and direction vector of descent, respectively, and j is the index of iteration.

A method for selecting the suitable σ_j and $\boldsymbol{v}_j$ in order to minimize $F(\boldsymbol{u})$ over the convex polytope U is discussed in [29], particularly as the algorithm is applied to color signal separation. In other words, the sequence $\{\boldsymbol{u}_j\}$ generated by (19) converges to $\boldsymbol{u}^*$, minimizing $F(\boldsymbol{u})$.

Ritter's superlinearly convergent method is proven as a computationally superior algorithm of finding the minimum points over a convex polytope. Unfortunately, this gradient-like algorithm sometimes will end up in a local minimum. The effect is that the algorithm appears to stop searching for the better minimum point. In the next section, we shall present one method for reducing the possibility of falling into a local minimum. That method is called simulated annealing because of its strong analogy to the physical annealing process done to metals and other substances.

IV. Linearly Constrained Simulated Annealing Approaches

Simulated annealing is a recently developed but promising approach to solving the combinatorial optimization problem. Kirkpatrik *et al.* [20] first recognized a strong analogy between physical annealing and solving large combinatorial optimization problems. In simulated annealing, the possible solutions of a combinatorial optimization problem are analogous to the states of a physical system, the cost of a given solution is analogous to the energy of a given state, and the control parameter T_e is analogous to the temperature of a heat bath. The annealing algorithm has been successfully applied to a variety of hard optimization problems in operational research [23], source coding [24], and image processing [25].

An instance of a combinatorial optimization problem [26] can be formalized as a pair (U, F), where the *solution state space* U denotes the finite set of possible solutions, and the *cost function* F is a mapping defined as $F : U \rightarrow \Re$. Hence, the goal of combinatorial minimization is to find the solution $\boldsymbol{u}_{\text{opt}} \in U$ such that $F(\boldsymbol{u}_{\text{opt}}) \leq F(\boldsymbol{u}_j)$ for all $\boldsymbol{u}_j \in U$. This goal may be achieved by using the following proposed simulated annealing algorithm.

Simulated annealing is a smart random search technique that is often more efficient than an exhaustive search yet more robust than gradient descent. Its behavior is controlled by an externally specified parameter, which is usually called effective temperature T_e, with the same unit as the cost function. When T_e is relatively large compared with the maximum value of the cost functon, simulated annealing explores the

entire solution space using a uniformly generated random perturbation, which is a transition from the current state (state $\boldsymbol{u}_j$ with cost $F(\boldsymbol{u}_j)$) to the proposed state (state $\hat{\boldsymbol{u}}$ with cost $F(\hat{\boldsymbol{u}})$) and with no preference for lower cost function. As T_e gets smaller, this undirected exploration changes. When T_e is small enough, simulated annealing becomes a descent algorithm. At intermediate values of T_e, the simulated annealing moderates its behavior between these two extremes: The transition is immediately accepted if $F(\hat{\boldsymbol{u}})$ is less than or equal to $F(\boldsymbol{u}_j)$; otherwise, it is accepted with a probability given by the Boltzman distribution $e^{(F(\boldsymbol{u}_j)-F(\hat{\boldsymbol{u}}))/T_e}$. By gradually decreasing T_e, the search is systematically concentrated into regions likely to contain a global minimum but still random enough to escape most local minimum. Corana [26] showed that the convergence to the global minima is provable with probability one sense.

In the following procedure, it is observed that there are two nested loops involved in the simulated annealing algorithm. The outer one will be terminated when T_e reaches the lowest effective temperature. In addition, the inner loop stops if the *equilibrium condition* [21] is satisfied. Moreover, Hajek [22] indicates that a cooling schedule process used to control the decrement of T_e would provide the conditions to ensure the convergence to the global minimum. The conditions are described as follows:

Lemma 2: A sequence of states generated according to the simulated annealing algorithm converges the global minimum for $F(\bullet)$ with probability one if and only if $T_e(t) = \frac{c}{\log(t+1)}$, where t is the index of the outer loop.

Unfortunately, the simulated annealing algotithm based on the cooling schedule of Lemma 2 converges to the desired solution extremely slowly. To overcome this difficulty, a pratical implementation is given by $T_e(t+1) = \vartheta T_e(t)$, where t is the index of the iteration for the outer loop, and the control parameter reduction factor ϑ is between 0.95 and 1.

Simulated Annealing Procedure

Step 0: [Initialization]

- $t = 0$.
- The initial effective temperature T_e is chosen so that the initial transition-acceptance ratio is very high [23]. Skiscim [21] suggested that the initial temperature can be chosen a value greater than the maximal cost function by an order of magnitude.
- Randomly generate an initial state $\boldsymbol{u}_0$ in the feasible solution state space U.

Step 1: [Start the effective temperature decremnt procedure]
WHILE (The lowest temperature is not reached)
BEGIN

$$\boldsymbol{u}_0 = \begin{cases} \boldsymbol{u}_0 & \text{if } t = 0 \\ \text{equilibrium point} & \text{if } t > 0 \end{cases} \tag{20}$$

and set $j = 0$

Step 2: [Start the procedure to reach the equilibrium point]
WHILE (*Equilibrium-Condition* is not satisfied)
BEGIN

Step 3: [State transition]
Generate a new state $\hat{\boldsymbol{u}}$ which belongs to U with a uniform random distribution
$\Delta F = F(\hat{\boldsymbol{u}}) - F(\boldsymbol{u}_j)$
IF $\Delta F < 0$ THEN Accept
$\boldsymbol{u}_{j+1} = \hat{\boldsymbol{u}}$
ELSE IF $e^{-\Delta F/T_e(M)} >$ RANDOM [0, 1]
THEN Accept
$\boldsymbol{u}_{j+1} = \boldsymbol{u}_j$
END IF
$j = j + 1$
END Do-Block of step 2

Step 4: [Cooling Schedule] $T_e(t+1) = \vartheta T_e(t)$
$t = t + 1$
END Do-Block of Step 1

It should be mentioned that the condition of reaching an equilibrium point at the nth iteration can be determined by the following equations:

$$\left|\frac{C_n - C_{n-1}}{C_n}\right| < \zeta \tag{21}$$

where ζ is a specific criterion, and C_n is given by

$$C_n = \sum_{j=1}^{n} \exp\{[AVE_{\text{accept}} - F(\boldsymbol{u}_j)]/T_e\}/n \tag{22}$$

where $\exp(\bullet)$ is a exponential function, and AVE_{accept} is the estimate of the average of the accepted cost functions and is given by

$$AVE_{\text{accept}} = \sum_{j=1}^{n_{\text{accept}}} F(\boldsymbol{u}_j)/n_{\text{accept}} \tag{23}$$

where n_{accept} is the number of acceptions.

The last problem remaining to be solved is a criterion that will terminate the outer loop and decide when the lowest temperature (or the global minimum) has been reached. Since the estimate of the lowest temperature is hard to find, the easiest way consists of stopping when the cost function has not significantly changed over a reasonable amount of temperature steps [21]. Specifically, the cost function value $F(\boldsymbol{u}_t)$ is recorded in the current equilibrium point $\boldsymbol{u}_t$ immediately before each temperature reduction from $T(t)$ to $T(t+1)$. When each cycle at temperature $T(t)$ is made, the search is stopped if

$$|F(\boldsymbol{u}_t) - F(\boldsymbol{u}_{t-\ell})| < \Delta, \quad \ell = 1, 2, \ldots, N_\Delta \tag{24}$$

where Δ is a termination criterion, and N_Δ is the number of temperature reductions to test for termination. Δ and N_Δ are the parameters that control the termination test. N_Δ has been kept equal to 5: a value found in exhaustive tests continuous optimization [28]. Δ is a classical "*termination test*" of optimization algorithms, and its value depends on the particular problem. In our opinion, a good pratice is to set Δ to a very low value (say $10^{-1} - 10^{-2}$) and observe at the temperature at which there is no further improvement: This will be the appropriate value of the lowest temperature.

A. Color Signal Separation by Simulated Annealing

To implement the simulated annealing algorithm incorporated with the problem of color signal separation, three things are required. First, we need to specify the solution state as an $(m+n)$ dimensional column vector $\boldsymbol{u} = [u_1, u_2, \ldots, u_{m+n}]^T = [\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T]^T$, where $u_i = \alpha_i$, $1 \le i \le m$ and $u_i = \beta_{i-m}$, $m+1 \le i \le m+n$. Second, the associated cost function is given by

$$F(\boldsymbol{u}) = \left\| \vec{I} - \sum_{i=1}^{m} \sum_{j=m+1}^{m+n} u_i u_{j-m} \vec{es}_{i,j-m} \right\|^2. \quad (25)$$

Finally, we need a method that can generate the feasible solution states belonging to the state space U with a uniform distribution. The state space U is given by

$$U = \{\boldsymbol{u} \mid \boldsymbol{A}\boldsymbol{u} \le \boldsymbol{b}\} = \hat{\Phi} \cap \Psi \quad (26)$$

where $\boldsymbol{b} = [(\vec{E}^+)^T, (\vec{S}^+)^T]^T$, and $\boldsymbol{A}$ is a $(4N+2) \times (m+n)$ matrix given by

$$\boldsymbol{A} = \begin{bmatrix} \Sigma & [0]_{(2N+2)\times n} \\ [0]_{(2N)\times m} & \Lambda \end{bmatrix}. \quad (27)$$

It has been shown in Lemma 1 that U forms a convex polytope. A process of generating feasible candidate states from U is identical to finding the interior points. The strategy is based on a hit-and-run algorithm [12] and introduced below.

B. Method of Generating the Interior Points of a Polytope

The hit-and-run algorithm is the attractive method used to generate a random sequence of interior points whose limiting distribution is uniform. The process of finding a new feasible interior point is done on the basis of searching in the direction of a random vector $\boldsymbol{v}_r$ from the current point (or feasible solution state) $\boldsymbol{u}_j$ and then choosing a new interior point $\hat{\boldsymbol{u}}$ uniformly on the line segment between the two hitpoints, i.e., the points where the line $\boldsymbol{u}_j + \gamma \boldsymbol{v}_r$ intersects the boundary of U and γ is any real scalar number and is known as the the line parameter. It is shown that the random direction vector $\boldsymbol{v}_r$ can be generated by both the hypersphere direction algorithm [12] and coordinate direction method [12]. In the hypersphere direction algorithm, the direction vector is drawn from a uniform distribution on a hypersphere. The coordinate direction method generates one of the unit coordinate vectors or their negation with equal probability as the direction vector. Berbee *et al.* [12] showed that the coordinate direction method is the computationally superior algorithm in generating the random direction vectors. The algorithm consists of the following steps:

Algorithm HAR (Hit–And–Run)

Step 0: [Initialization]

i) Given an interior point $\boldsymbol{u}_j$.

Step 1: [Direction Vector]

Generate a direction vector $\boldsymbol{v}_r$ with equal probablity from one of $(m+n)$ coordinate vectors and their opposite vectors

Step 2: [Line Parameters]

Determine:

i) $\gamma_i = (b_i - \boldsymbol{a}_i^T \boldsymbol{u}_j)/(\boldsymbol{a}_i^T \boldsymbol{v}_r)$, $i = 1, 2, \ldots, 4N+2$

ii) $\gamma^+ = \min_{1 \le i \le 4N+2}\{\gamma_i \mid \gamma_i > 0\}$

iii) $\gamma^- = \max_{1 \le i \le 4N+2}\{\gamma_i \mid \gamma_i < 0\}$, where $\boldsymbol{a}_i^T$ and b_i is the ith row and ith entry of $\boldsymbol{A}$ and $\boldsymbol{b}$, respectively, and γ_i is the line parameter for the ith constraint $\boldsymbol{a}_i^T \boldsymbol{u} \le b_i$.

Step 3: [New Interior Point]

i) Generate θ from a uniform distribution on [0, 1]

ii) Let $\hat{\boldsymbol{u}} = \boldsymbol{u}_j + [\gamma^- + \theta(\gamma^+ - \gamma^-)]\boldsymbol{v}_r$

Step 4: [Output and Terminate]

It is known that the computational complexity of calculations involved in simulated annealing algorithm is in proportion to the exponential of the dimensionality of its associated state space. Hence, the convergence of annealing algorithm becomes extremely slow while the dimensionality grows larger. To tackle this undesired effect, one of the possible methods is to divide the large-scale state space into several small-scale state spaces. Each of them can be performed in an efficient way individually. As a consequence, the subresults derived from these small-scale problems can then be combined in order to yield a result that is identical to that of the original problem. To achieve this goal, we adopt a method proposed by Golub and Pereyra [11]. They showed that the nonlinear least squares problem can be reduced to be a small-scale nonlinear least squares problem when their variables are separable. More detail about this method will be discussed in the next section.

V. A Variable-Separable Nonlinear Least Squares Approach

A general nonlinear least squares problem is conducted to apply the least squares fit techniques to the nonlinear models $\eta(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$, where $\boldsymbol{\alpha} \in \hat{\Phi} \subseteq \Re^n$ and $\boldsymbol{\beta} \in \Psi \subseteq \Re^m$ are unknown parameter vectors to be determined, and λ is the independent variable. More precisely, given the data $(\lambda_k, I(\lambda_k))$, $k = 1, 2, \ldots, N, N \ge m+n$, the task is to find the values of the parameters α and β that minimize the nonlinear objective functional

$$\begin{aligned} d(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \sum_{k=1}^{N} (I(\lambda_k) - \eta(\boldsymbol{\alpha}, \boldsymbol{\beta}; \lambda_k))^2 \\ &= \|\vec{I} - \vec{\eta}(\boldsymbol{\alpha}, \boldsymbol{\beta})\|^2 \end{aligned} \quad (28)$$

where $\vec{I} = [I(\lambda_1), I(\lambda_2), \ldots, I(\lambda_N)]^T$, and $\vec{\eta}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = [\eta(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda_1), \eta(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda_2), \ldots, \eta(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda_N)]^T$.

The nonlinear least squares is said to be variable separable if the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ form two completely disjoint sets, and the nonlinear model $\eta(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda)$ is expressed as

$$\eta(\boldsymbol{\alpha}, \boldsymbol{\beta}; \lambda) = \sum_{j=1}^{n} \beta_j \varphi_j(\boldsymbol{\alpha}; \lambda) \quad (29)$$

where $\varphi_j(\boldsymbol{\alpha}; \lambda)$ is the jth associated nonlinear (or linear) functional of $\boldsymbol{\alpha}$ and λ, or equivalently

$$\vec{\eta}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = M(\boldsymbol{\alpha})\boldsymbol{\beta} \quad (30)$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_n]^T$ and $M(\boldsymbol{\alpha})$ is an $N \times n$ matrix function of α with its (k, j)-entry $(M(\boldsymbol{\alpha}))_{kj} = \varphi_j(\boldsymbol{\alpha}; \lambda_k); 1 \leq k \leq N$, and $1 \leq j \leq n$. Substituting (30) into (28), it becomes

$$d(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \|\vec{I} - M(\boldsymbol{\alpha})\boldsymbol{\beta}\|^2. \tag{31}$$

Comparing (29) and (4), it can be found that our nonlinear least squares formulation for color signal separation is indeed a variable separable problem. It yields

$$(M(\boldsymbol{\alpha}))_{kj} = \sum_{i=1}^{m} \alpha_i (e_i(\lambda_k) s_j(\lambda_k)) \tag{32}$$

where $1 \leq k \leq N$ and $1 \leq j \leq n$, or equivalently

$$M(\boldsymbol{\alpha}) = E(\boldsymbol{\alpha}) S^0 \tag{33}$$

where $E(\boldsymbol{\alpha}) = \text{diag}[\sum_{i=1}^{m} \alpha_i e_i(\lambda_1), \sum_{i=1}^{m} \alpha_i e_i(\lambda_2), \ldots \sum_{i=1}^{m} \alpha_i e_i(\lambda_N)]$ and $S_0 = [\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_N]^T$.

Lemma 3: The $N \times n$ matrix $M(\boldsymbol{\alpha})$ has constant rank $r \leq \min(N, n)$ for $\boldsymbol{\alpha} \in \Omega (= \hat{\Phi}|_\Psi)$, where $r = \text{rank}(S^0)$ and $\hat{\Phi}|_\Psi$ is a set of feasible vectors $\boldsymbol{\alpha}$ belonging to $\hat{\Phi}$ subject to the constraint region Ψ.

Since $E(\alpha)$ is a diagonal matrix with nonzero entries on its main diagonal and of full rank for $\boldsymbol{\alpha} \in \Omega$, it is easily proven that the rank of $M(\boldsymbol{\alpha})$ is identical to that of S^0.

In order to perform the variable separable problem, a new objective functional of $\boldsymbol{\alpha}$ is required and given by

$$d_2(\boldsymbol{\alpha}) = \|\vec{I} - M(\boldsymbol{\alpha}) M^+(\boldsymbol{\alpha}) \vec{I}\|^2 \tag{34}$$

where $M^+(\boldsymbol{\alpha})$ is a pseudoinverse of $M(\boldsymbol{\alpha})$.

This new functional is called the variable projection functional. For any given $\boldsymbol{\alpha}$, the minimal least squares solution to $M(\boldsymbol{\alpha})\boldsymbol{\beta} \cong \vec{I}$ is given by

$$\boldsymbol{\beta}^*(\boldsymbol{\alpha}) = M^+(\boldsymbol{\alpha}) \vec{I}. \tag{35}$$

Substituting (35) into (31), we obtain

$$\begin{aligned} &\min_{\boldsymbol{\beta} \in \Psi} d(\boldsymbol{\alpha}, \boldsymbol{\beta}) = d(\boldsymbol{\alpha}, \boldsymbol{\beta}^*(\boldsymbol{\alpha})) \\ &\quad = \|\vec{I} - M(\boldsymbol{\alpha}) M^+(\boldsymbol{\alpha}) \vec{I}\|^2 = d_2(\boldsymbol{\alpha}). \end{aligned} \tag{36}$$

It turns out that the variable projection functional equals the minimum value of $d(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ over Ψ. Furthermore, Golub and Pereyra [11] showed that once optimal parameter vector $\boldsymbol{\alpha}^*$ has been obtained by minimizing (34), then the parameter vector $\boldsymbol{\beta}^*$ is obtained as a solution of $M(\boldsymbol{\alpha}^*)\boldsymbol{\beta} \cong \vec{I}$. The relationship between critical points or minimal points of the original objective functional $d(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and those obtained from the functional $d_2(\boldsymbol{\alpha})$ is described below.

Theorem 2 [11]: Let $d(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $d_2(\boldsymbol{\alpha})$ be defined as above. $M(\boldsymbol{\alpha})$ is assumed to have constant rank over Ω.

a) If $\boldsymbol{\alpha}^*$ is a critical point (or a global minimizer in Ω) of $d_2(\boldsymbol{\alpha})$ and

$$\boldsymbol{\beta}^* = \boldsymbol{\beta}^*(\boldsymbol{\alpha}^*) = M^+(\boldsymbol{\alpha}^*) \vec{I} \tag{37}$$

then $\boldsymbol{u}^* = [\boldsymbol{\alpha}^{*T}, \boldsymbol{\beta}^{*T}]^T$ is a critical point (or a global minimizer) for $\boldsymbol{u} \in U$ and $d(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = d_2(\boldsymbol{\alpha}^*)$, where $\Omega = \hat{\Phi}|_\Psi$ and U is defined in (26).

b) Conversely, if $\boldsymbol{u}^*$ is a global minimizer of $d(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for $\boldsymbol{u} \in U$, then $\boldsymbol{\alpha}^*$ is a global minimizer in Ω, and $d_2(\boldsymbol{\alpha}^*) = d(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Furthermore, if there is a unique $\boldsymbol{\alpha}^*$ among minimizing pairs of $d(\boldsymbol{\alpha}, \boldsymbol{\beta})$, then $\boldsymbol{\alpha}^*$ must satisfy (35).

Let $P_{M(\boldsymbol{\alpha})}$ and $P^{\perp}_{M(\boldsymbol{\alpha})}$ be the orthogonal projection operator on the linear subspace spanned by the columns of the matrix $M(\boldsymbol{\alpha})$ and the projector on the orthogonal complement of the column space, respectively. Expressing

$$P_{M(\boldsymbol{\alpha})} = M(\boldsymbol{\alpha}) M^{\perp}(\boldsymbol{\alpha}) \tag{38}$$

and

$$P^{\perp}_{M(\boldsymbol{\alpha})} = \boldsymbol{I}_N - P_{M(\boldsymbol{\alpha})} \tag{39}$$

where $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix, and substituting (38) into (34), we obtain

$$d_2(\boldsymbol{\alpha}) = \|P^{\perp}_{M(\boldsymbol{\alpha})} \vec{I}\|^2 \tag{40}$$

Moreover, one can avoid evaluating the nonlinear functional $d_2(\boldsymbol{\alpha})$ from $P^{\perp}_{M(\boldsymbol{\alpha})}$ directly in order to reduce the computational burden of calculations involved in determining $P^{\perp}_{M(\boldsymbol{\alpha})}$ by using the complete orthogonal decomposition on both $M(\boldsymbol{\alpha})$ and $P^{\perp}_{M(\boldsymbol{\alpha})}$ [10].

By applying the simulated annealing technique to find the minimum point of $d_2(\boldsymbol{\alpha})$ over the state space $\Omega = \hat{\Phi}|_\Psi$, the suggested efficient calculation of $d_2(\boldsymbol{\alpha})$ is required to speed up the convergence. Once the global minimum point $\boldsymbol{\alpha}^*$ is found, then $\boldsymbol{\beta}^*$ is obtained by replacing $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha}^*$ in (35). Apart from the linearly constrained simulated annealing discussed in the above section, the state space $\Omega = \hat{\Phi}|_\Psi$ has nonlinear constraints. One cannot generate the interior points belonging to $\Omega = \hat{\Phi}|_\Psi$ by the original hit-and-run algorithm directly. Before introducing the modifications, the nonlinear constraints in $\Omega = \hat{\Phi}|_\Psi$ are derived from Ψ and given by

$$0 \leq \vec{s}_i^T \boldsymbol{\beta}^*(\boldsymbol{\alpha}) (= S(\lambda_i)) \leq 1 \tag{41}$$

or equivalently

$$0 \leq \vec{s}_i^T M^+(\boldsymbol{\alpha}) \vec{I} \leq 1 \tag{42}$$

where $\vec{s}_i^T = [s_1(\lambda_i), s_2(\lambda_i), \ldots, s_n(\lambda_i)]$, and $1 \leq i \leq N$.

Mathematically, those nonlinear constraints can be rephrased as a set of $2N$ inequality constraints

$$f_k(\boldsymbol{\alpha}) = \vec{s}_k^T M^+(\boldsymbol{\alpha}) \vec{I} - 1 \leq 0 \quad \text{if } 1 \leq k \leq N \tag{43.a}$$

or

$$f_k(\boldsymbol{\alpha}) = -\vec{s}_{k-N}^T M^+(\boldsymbol{\alpha}) \vec{I} \leq 0 \quad \text{if } N+1 \leq k \leq 2N \tag{43.b}$$

where $f_k(\boldsymbol{\alpha})$ is the kth nonlinear constraint function from an m-dimensional Euclidean space $\Re^m$ into a set of real number. In the case of nonlinear constraint functions, only Step 2 of the hit-and-run algorithm changes.

A. Method of Finding the Line Parameters of Nonlinear Constraint Functions

The strategy to find the line parameter of a prescribed nonlinear constraint function is based on determining the intersection between the line $\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r$ and the curve $f_k(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}_j$ and $\boldsymbol{v}_r$ are the given current point and direction vector, respectively, and ς is the line parameter for the nonlinear constraint. An alternative way of finding this intersection effectively is to define a new parametric function in terms of the nonlinear constraint functions

$$g_k(\varsigma) = f_k(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r) \tag{44}$$

where $g_k(\bullet)$ is a real-valued function from $\Re$ to $\Re$.

Then, the Newton's method [17] could be applied to find the root ς_k of $g_k(\varsigma)$, i.e, $g_k(\varsigma)|_{\varsigma=\varsigma_k} = 0$. The Newton's iteration is

$$\varsigma^{n+1} = \varsigma^n - [g'(\varsigma^n)]^{-1} g_k(\varsigma^n) \tag{45}$$

where $g'_k(\varsigma^n) = \frac{dg_k(\varsigma)}{d\varsigma}|_{\varsigma=\varsigma^n}$

As shown in (43.a) and (43.b), there are two expressions for $g_k(\varsigma)$. First, we would like to derive the explicit expression for the first order derivative of the parametric function of (43.a) with respect to ς. The parametric function can be expressed as

$$g_k(\varsigma) = \vec{s}_k^T M^+(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r)\vec{I} - 1. \tag{46}$$

Its derivative is given by

$$g'_k(\varsigma) = \vec{s}_k^T \left(\frac{d}{d\varsigma} M^+(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r)\right)\vec{I} \tag{47}$$

where $1 \le k \le N$.

In order to obtain the expression for the first-order derivative of a specific pseudoinverse with respect to ς, one should follow from the following results. Let $\boldsymbol{B} = \boldsymbol{A} + d\boldsymbol{A}$, where $d\boldsymbol{A}$ is an arbitrary incremental matrix. Wedin [18], [19] has shown that

$$\begin{aligned} &\boldsymbol{B}^+ - \boldsymbol{A}^+ \\ &\quad = -\boldsymbol{B}^+ d\boldsymbol{A}\boldsymbol{A}^+ + \hat{P}_{\boldsymbol{B}}^{\perp} d\boldsymbol{A}^T \boldsymbol{A}^{+^T}\boldsymbol{A}^+ + \boldsymbol{B}^+\boldsymbol{B}^{+^T} d\boldsymbol{A}^T \boldsymbol{P}_{\boldsymbol{A}}^{\perp} \end{aligned} \tag{48}$$

where $\hat{P}_{\boldsymbol{B}}^{\perp} = \boldsymbol{I}_n - \boldsymbol{B}^+\boldsymbol{B}$ is defined as the projector on the orthogonal complement of the row space of $\boldsymbol{B}$, and $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix.

Let $\boldsymbol{A}(\varsigma) = M(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r)$ and $\boldsymbol{B} = \boldsymbol{A}(\varsigma + d\varsigma)$, where $d\varsigma$ is a small scalar increment. From (33), it can be shown that $M(\boldsymbol{x})$ is a linear mapping satisfying

$$M(\boldsymbol{x} + \xi \boldsymbol{y}) = M(\boldsymbol{x}) + \xi M(\boldsymbol{y}) \tag{49}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are arbitrary vectors, and ξ is any scalar number. Using this, we obtain

$$\begin{aligned} \boldsymbol{B} = \boldsymbol{A}(\varsigma + d\varsigma) &= M\{(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r) + d\varsigma \boldsymbol{v}_r\} \\ &= M(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}) + d\varsigma M(\boldsymbol{v}_r) \\ &= \boldsymbol{A}(\varsigma) + d\boldsymbol{A} \end{aligned} \tag{50}$$

where $d\boldsymbol{A} = d\varsigma M(\boldsymbol{v}_r)$.

Finally, the first-order derivative of $\boldsymbol{A}(\varsigma)$ with respect to ς can then be derived by letting $d\varsigma \to 0$. It yields

$$\begin{aligned} \frac{d}{d\varsigma} M^+(\boldsymbol{\alpha}_j + \varsigma \boldsymbol{v}_r) &= \frac{d}{d\varsigma}\boldsymbol{A}^+(\varsigma) \\ &= -\boldsymbol{A}^+ M(\boldsymbol{v}_r)\boldsymbol{A}^+ + \hat{P}_{\boldsymbol{A}}^{\perp} M^T(\boldsymbol{v}_r)\boldsymbol{A}^{+^T}\boldsymbol{A}^+ \\ &\quad + \boldsymbol{A}^+\boldsymbol{A}^{+^T} M^T(\boldsymbol{v}_r)\boldsymbol{P}_{\boldsymbol{A}}^{\perp}. \end{aligned} \tag{51}$$

Similiarly, the first-order derivative of the other expression becomes

$$g'_k(\varsigma) = -\vec{s}_{k-N}^T \frac{dA^+(\varsigma)}{d\varsigma}\vec{I} \tag{52}$$

where $N + 1 \le k \le 2N$ and $\frac{dA^+(\varsigma)}{d\varsigma}$ is defined in (51).

Hence, the roots of $g_k(\varsigma)$ for $N+1 \le k \le 2N$ can be found by the Newton's iteration based on the first-order derivative of (52).

Following step 2 of algorithm HAR, two essential parameters are made for the nonlinear constraint functions and are given by

$$\varsigma^+ = \min\{\varsigma_k \mid \varsigma_k > 0, g_k(\varsigma_k) = 0\} \tag{53.a}$$

and

$$\varsigma^- = \max\{\varsigma_k \mid \varsigma_k < 0, g_k(\varsigma_k) = 0\}. \tag{53.b}$$

To accomplish the generation of the interior points belonging to $\hat{\Phi}|_{\Psi}$, both the parameters ς^+ and ς^- should be incorporated with the parameters γ^+ and γ^- derived from the linear constraints $\hat{\Phi} = \{\boldsymbol{\Sigma}\alpha \le \vec{E}^+\}$. The parameters that satisfy both the linear and nonlinear constraints are defined as

$$\hat{\gamma}^+ = \min(\gamma^+, \varsigma^+) \tag{54.a}$$

and

$$\hat{\gamma}^- = \max(\gamma^-, \varsigma^-). \tag{54.b}$$

It can be shown that these new parameters can make the interior points fall in the region $\hat{\Phi}|_{\Psi}$.

B. Fast Simulated Annealing

The cooling schedule is critical to the performance of the annealing algorithm. For a given random process, cooling at too fast at rate will likely "freeze" in a nonglobal minimum. Cooling at a too slow rate, while reaching the desired minimum, is a waste of computational resources. Since the classical simulated annealing is derived from the Boltzman random process, it has been proven [22] that the cooling schedule is inversely proportional to the logarithm of t in order to guarantee convergence to the global minimum, where t is the index of the outer loop of annealing algorithm. This relatively slow convergence is due to the bounded variance of the Boltzman process that constrains the neighborhood of the successive samples. On the other hand, Szu [29] proposed a method based on Cauchy random process whose variance is infinite. The Cauchy distribution has the same general shape as the Boltzman distribution but does not fall off as sharply at large energies. The implication is that the Cauchy annealing may occasionally take a rather large jump out of

a local minimum. The advantage of this approach is that the global minimum can be reached with much shorter annealing schedule. For the Cauchy annealing, the temperature should follow the inverse of t:

$$T(t) = \frac{T_0}{1+t} \tag{55}$$

where T_0 is the initial temperature, and t denotes the index (or "time") of outer loop.

VI. Simulation Results

To verify the effectiveness of the proposed optimization techniques and compare them with the Ho *et al.* data [3], a number of test samples used in their examples are particularly considered in our simulation. Ho *et al.* tested their algorithm by combining each of the 370 available surface reflectance functions recorded by Krinov [3] with each of the standard daylight illumination spectra given by Judd [5]. The finite-dimensional models are based on the first three of Cohen's vectors [6] as basis functions for surface spectral reflectance and the first three of Judd's vectors [5] as basis functions for spectral power distribution of ambient light. According to Theorem 1, it is shown that any separation algorithm is valid when both the color signal and illuminant are assumed to be normalized. As an example, Fig. 1 illustrates a normalized color signal that has been constructed from the surface spectral reflectance of Krinov's sample 54 multiplied by the normalized Judd's spectrum for 4800° K correlated color temperature daylight. The Krinov's sample 54 is a standard image of "river valley meadow." This normalized color signal is sampled at 10-nm intervals from 400 to 650 nm. After performing any separation algorithm on the given sampled color signal, we would like to evaluate the performance of the algorithm by examining the root mean-square error formed by the square of the difference between the actual color signal and model's approximation of it in a normalization sense. This error is identical to the square root of (5), i.e., $\epsilon_{ES} \equiv \sqrt{d(\boldsymbol{\alpha}, \boldsymbol{\beta})}$. While reaching the minimum point, $\epsilon^*_{ES} \equiv \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sqrt{d(\boldsymbol{\alpha}, \boldsymbol{\beta})} = \sqrt{d(\boldsymbol{\alpha}, \boldsymbol{\beta})|_{\boldsymbol{\alpha}=\boldsymbol{x},\boldsymbol{\beta}=\boldsymbol{y}}}$. Additionally, we still need to measure the errors in the individual fits to the illumination and surface spectral reflectance based on the optimal values $\boldsymbol{x}$ and $\boldsymbol{y}$. Both the errors are assumed to be in a relative and normalization sense and defined as

$$\epsilon^*_E = \frac{\|\sum_{i=1}^3 x_i \vec{e}_i - \vec{E}\|}{\|\vec{E}\|}\Big|_{x_1=1,\alpha_1=1} \tag{56}$$

and

$$\epsilon^*_S = \frac{\|\sum_{j=1}^3 y_i \vec{s}_j - \vec{S}\|}{\|\vec{S}\|}. \tag{57}$$

Practically, the variable α_1 in our simulation is always set to be constant and identical to one. Meanwhile, this does not have any influence on the optimization formulation discussed above. Hence, the dimensionality of the entire solution space is reduced to only five. For the variable-separable formulation, the dimensionality of solution space equals two. It is believed that the computational complexity can then be decreased significantly. Two typical cases of input color signals shown in Fig. 1 and Fig. 4 are conducted to demonstrate the performance of those optimization techniques and are described below. The color signal shown in Fig. 4 and of particular interest is constructed from Krinov reflectance sample 162 illuminated by 4800°K daylight. Ho *et al.* [3]

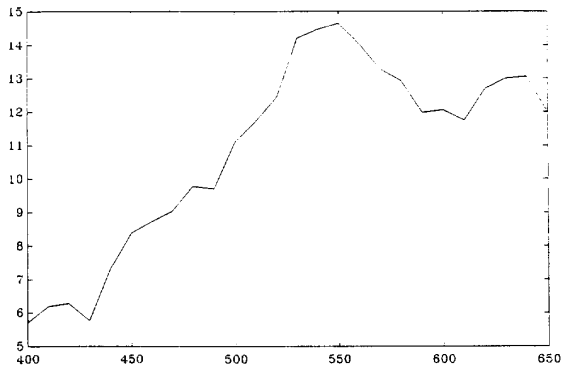


Fig. 1. Normalized color signal from product of surface spectral reflectance of river valley with meadows (Krinov spectrum 54) and Judd *et al.* spectrum for 4800°K daylight.

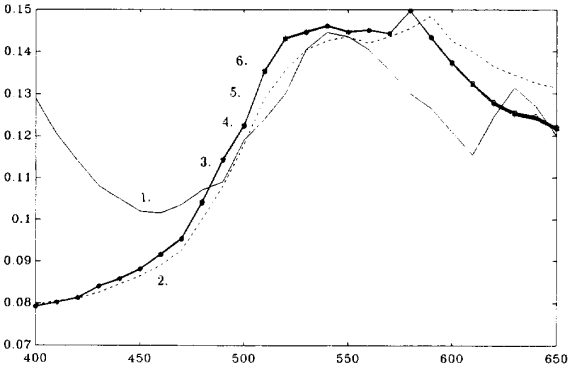


Fig. 2. Original Krinov reflectance spectrum 54 and results of Ho, SCM, LCSA, VSM, and FVSM separation algorithms applied to the color signal of Fig. 1 are denoted by "1," "2," "3," "4," "5," and "6," respectively.

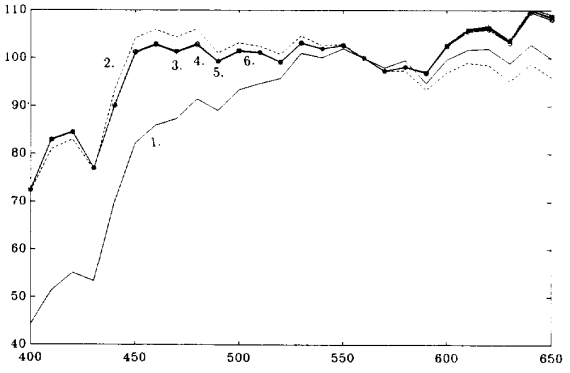


Fig. 3. Original Judd *et al.* normalized spectral power distribution of 4800°K, and results of Ho, SCM, LCSA, VSM, and FVSM separation algorithms applied to the color signal of Fig. 1 are denoted by "1," "2," "3," "4," "5," and "6," respectively.

TABLE I

COMPARISON OF SIMULATION RESULTS OF HO, SCM, LCSA, VSM, AND FVSM SEPARATION ALGORITHMS. 1) NORMAL CASE, 2) WORST CASE. $\boldsymbol{x} = [1, x_2, x_3]^T$ AND $\boldsymbol{y} = [y_1, y_2, y_3]^T$ ARE THE OPTIMAL COEFFICIENTS FOR $\vec{E}$ AND $\vec{S}$, RESPECTIVELY. T: COMPUTATION TIME EXECUTED ON IBM PC-486. "-": NOT AVAILABLE.

		ϵ^*_{ES}	ϵ^*_E	ϵ^*_S	x_2	x_3	y_1	y_2	y_3	$T(sec)$
Ho	1	4.8836	15.94%	15.2%	-0.4973	1.2074	0.0283	0.0122	-0.0140	-
	2	-	-	47.4%	-	-	0.022	0.0577	-0.037	-
SCM	1	2.9911	15.31%	14.54%	-0.4390	3.0114	0.0248	0.0282	-0.0093	4.715×10^2
	2	12.8755	85.47%	47.70%	-0.3389	29.6217	0.0163	0.0818	-0.0472	6.638×10^2
LCSA	1	2.9914	15.48%	14.47%	-0.4362	3.0730	0.0248	0.0282	-0.0095	4.8181×10^4
	2	11.3924	35.61%	46.72%	-1.2519	-10.4720	0.0369	0.0290	0.0600	3.7977×10^4
VSM	1	2.9909	15.41%	14.47%	-0.4412	2.9816	0.0248	0.0281	-0.0092	7.7823×10^3
	2	11.3924	35.61%	46.72%	-1.2519	-10.4720	0.0369	0.0292	0.0600	5.4329×10^3
FVSM	1	2.9909	15.47%	14.52%	-0.44027	2.9275	0.0248	0.0281	-0.0091	4.0546×10^3
	2	11.3924	35.58%	46.70%	-1.2571	-10.4789	0.0369	0.0290	0.0599	2.6216×10^3

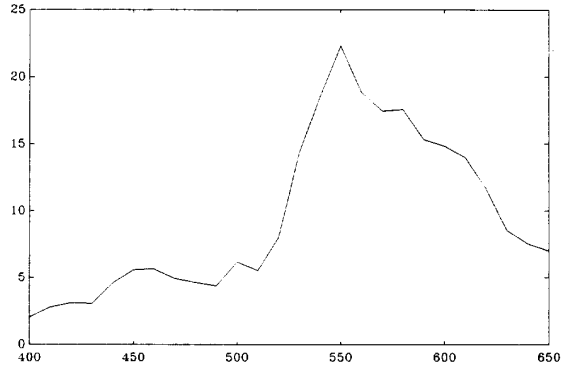


Fig. 4. Normalized color signal from product of surface spectral reflectance of Krinov reflectance 162 and Judd *et al.* spectrum for 4800°K daylight.

showed that it is the worst case for their separation algorithm.

The superlinearly convergent algorithm is applied to the color signal separation with initial point $\boldsymbol{u} = \mathbf{0}$. The parameters involved in the algorithm are set to be $\chi = 0.6$, $\gamma = 10^{-3}$, $\gamma' = 0.5 \times 10^{-3}$, $m' = 10^2$, $m^0 = 10^4$, $\tau = 10^4$, and $\delta = 0.24$. In addition, the related sequences are assigned to be $m_j = m' - \frac{2}{j}$, $\gamma_j = \gamma + \frac{2}{j}$, $\gamma'_j = \gamma' + \frac{2}{j}$. The algorithm applied to the normalized color signal shown in Fig. 1 proceeds iterately and stops at a minimum point $\boldsymbol{u}^* = (\alpha_2^*, \alpha_3^*, \beta_1^*, \beta_2^*, \beta_3^*)^T = (-0.4390, 3.0114, 0.0248, 0.0282, -0.0093)^T$ with $\epsilon^*_{ES} = 2.9911$. As shown in Table I, the algorithm results in an minimum point that is almost identical to that of the other three methods and has the fastest convergence rate. According to Lemma 2, this point is regarded as the global minimum, but it may fall into a local minimum point $(= (-0.3389, 29.6217, 0.0163, 0.0818, -0.0472)^T)$ and produces the larger error $\epsilon^*_{ES} = 12.8755$ when the input normalized color signal is the second case shown in Fig. 4. Moreover, it should be mentioned that the superlinearly convergent algorithm will become unstable and be trapped in a undesired point when those parameters and sequences involved in the algorithm are not chosen properly.

The remaining three methods are derived from a stochastic annealing-like optimization technique in statistical mechanics. The annealing algorithm has been shown as the most promising global minimizer applied to a variety of hard optimization problems successfully. In constrast to the superlinearly convergent method, the global convergence of the annealing algorithm is less sensitive to variation in parameters and is always guaranteed. The first approach uses the annealing algorithm to find the global minimum solutions of the typical two cases over a 5-D convex polytope since $\alpha_1 = 1$. The initial temperature is chosen to be a very high value that is greater than the maximal value of $d(\boldsymbol{\alpha}, \boldsymbol{\beta})(\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} d(\boldsymbol{\alpha}, \boldsymbol{\beta}) = O(10^5))$ by an order of magnitude, i.e., $T_e(0) = 10^6$ and reduction factor $\vartheta = 0.98$, respectively. The algorithm will be terminated at a very low value of $T_e (= 1.05 \times 10^{-1})$ and yield the global minimum points $[-0.4362, 3.0730, 0.0248, 0.0282, -0.0095]^T$ with $\epsilon^*_{ES} = 2.9914$ and $[-1.2519, -10.4720, 0.0369, 0.0290, 0.0600]^T$ with $\epsilon^*_{ES} = 11.3924$ for both cases, but the convergence of the algorithm is extremely slow. The computation times for both two cases executed on an IBM PC-486 are proportional to 10^4 s.

To eliminate this unfairly slow convergence, a cost-effective and robust approach based on the concept of variable-separable formulation is proposed to reduce the dimensionality of its solution space and improve its efficiency. The simulated annealing is applied to finding the minimum value of variable project functional $d_2(\boldsymbol{\alpha})$ over $\Omega (= \hat{\Phi}|_\Psi)$ of dimension two. The initial temperature and reduction factors are set to be 10^3 and 0.98, respectively, since $\max_\alpha d_2(\boldsymbol{\alpha}) = O(10^2)$. It is found in Table I that the computation times for both cases have been improved by an order of magnitude and in proportion to 10^3 s. Moreover, Table I shows that the minimum points obtained from variable-separable formulation for both cases are almost identical to that of the first approach but with a fast convergence rate. To improve the computational efficiency further, one may use fast simulated annealing instead of the classical annealing algorithm to find the global minimum point of the variable projection functional. As shown in Table I, the convergence rate of the fast simulated annealing for both cases is more fast than twice that of the classical annealing algorithm, but it results in the same solution as the classical one. On performing the generation of interior points belonging to Ω during the execution of the annealing algorithm, the process of finding the line parameters of the nonlinear constraint functions defined in (43.a) and (43.b) is worthy of being observed. For the case of input color signal shown in Fig. 1, we have found that the nonlinear constraint functions defined in (43.a) are strictly less than zero for $\boldsymbol{\alpha}_1 = 1$ and $\boldsymbol{\alpha} = [\alpha_2, \alpha_3]^T \in \Re^2$. This implies that their associated

inequality constraints are always satisfied. Therefore, only the nonlinear constraints defined in (43.b) are considered. To illustrate its function, we randomly choose the 438th iteration as an example. There are 26 nonlinear constraint functions at the 438th iteration of the annealing algorithm. Additionaly, the interior point obtained from previous iteration and the direction vector generated by coordinate direction method are $\boldsymbol{\alpha}_{437} = [1.9312, 3.6619]^T$ and $\boldsymbol{v}_r = [1, 0]^T$, respectively. Each line parameter of its associated nonlinear constraint function can be found by Newton's iteration. Fig. 6 shows time evolution of six typical functions. One may note that the convergence of each curve is very fast. The slowest one is the fourth curve, which reaches the horizontal axis at the tenth iteration. Its associated line parameter ς_4 is equal to -8449.8. After the line parameters for the 26 nonlinear constraints have been determined, ς^+ and ς^- become 37.2 and -5.9, respectively. Similarly, the parameters γ^+ and γ^- for linear constraints are found to be 4.4870 and -3.8922, respectively. Hence, $\hat{\gamma}^+ = 4.4870$, and $\hat{\gamma}^- = -3.8922$. It can be easily shown that $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_{437} + [\hat{\gamma^+} + \theta(\hat{\gamma^+} - \hat{\gamma^-})]\boldsymbol{v}_r$ is the new random interior point belonging to Ω, where $\theta \in [0, 1]$. As a result, the minimum point $\boldsymbol{\alpha}^*$ has been obtained by performing the annealing algorithm on $d_2(\boldsymbol{\alpha})$ for first case and found to be $\boldsymbol{\alpha}^* = [-0.4412, 2.9816]^T$. From Theorem 2, the other set of optimal variables $\boldsymbol{\beta}^* = [\beta_1^*, \beta_2^*, \beta_3^*]^T$ can then be determined as $\boldsymbol{\beta}^* = M^+(\alpha^*)\vec{I} = [0.0248, 0.0281, -0.0092]^T$.

Fig. 2 shows the original surface reflectance spectrum given by Krinov's sample 54, along with the results from Ho's and our optimization-based separation algorithms. The separation errors ϵ_S^* defined in (57) for Ho, the superlinearly convergent method (SCM), linearly constrained simulated annealing (LCSA), the variable-separable method (VSM), and the variable-separable method based on the fast simulated annealing (FVSM) are 15.2, 14.54, 14.47, 14.47, and 14.52%, respectively. These four optimization methods result in almost the same separation error, which is better than that of Ho. For the worst case, Fig. 5 shows Krinov's reflectance sample 162 spectrum along with the results from Ho's and our separation algorithms with the separation errors 47.4% for Ho, 47.70% for SCM, 46.72% for LCSA, 46.72% for VSM, and 46.70% for FVSM, respectively. It is seen that these three annealing-based separation algorithms yield the results better than those obtained from Ho and SCM.

Fig. 3 shows Judd's 4800°K daylight spectrum derived from the normalized color signal shown in Fig. 1, along with the normalized illuminations calculated by Ho's and our optimization separation algorithms. The normalized illumination errors ϵ_E^* defined in (56) for Ho, SCM, LSCA, VSM, and FVSM are 15.94, 15.31, 15.48, 15.41 and 15.47%, respectively. As seen in Ho's paper, their illumination error is not in a normalization sense. Here, we convert it (12.8%) into a normalized one (15.94%). One may find that our results are slightly better than Ho's results.

VII. Conclusion

This paper presents the development of a constrained nonlinear least-squares error formulation for recovering both the physically realiable illumination and surface reflectance from a given color signal. For ease of performing color-signal separation, the formulation should be parameterized according to the finite-dimensional linear approximate models. Thus, the process of separating a color signal is then reformulated as finding the optimal coefficients involved in those linear models subject to a number of feasibility constraints. The optimal solution can be carrier out by the superlinearly convergent method in the fastest convergence rate, but Table I shows that the superlinearly convergent method may be trapped in a local minimum. Hence, a promising global minimizer known as the annealing algorithm is presented and reduces the possibility of falling into the local minimum. Unfortunately, its convergence is extremely slow. Here, we have developed a cost-effective procedure based on the concept of variable separability, which leads to the same global minimum and has a reasonable convergence rate. The superiority of variable-separability comes from the fact that it reduces the dimensionality of solution space. Actually, the computational efficiency is further improved by the fast simulated annealing incorporated with the variable-separable method. Computer simulation shows that the computation time of the cost-effective method is between those derived from the superlinearly convergence method and the original annealing algorithm.

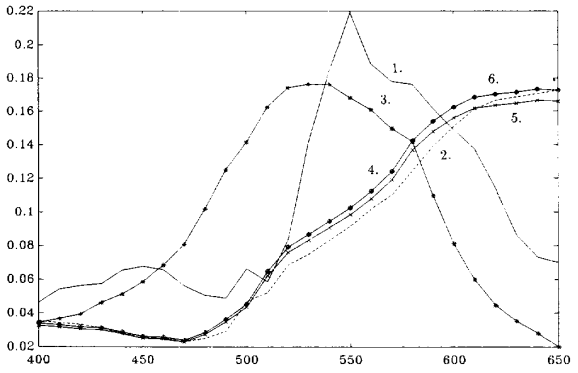


Fig. 5. Original Krinov reflectance spectrum 162 and results of Ho, SCM, LCSA, VSM, and FVSM separation algorithms applied to the color signal of Fig. 4 are denoted by "1," "2," "3," "4," "5," and "6," respectively.

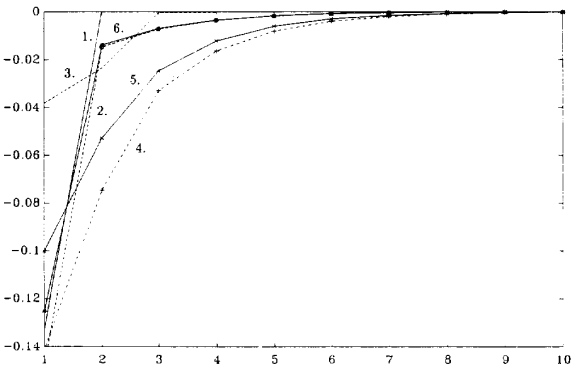


Fig. 6. Time evolution of six typical nonlinear constraint functions are denoted by "1," "2," "3," "4," "5," and "6," respectively.

APPENDIX A
COMPUTATION OF THE ∇F_j

The ith component of the gradient $\nabla F(\boldsymbol{u})$ can be evaluated by taking the first-order partial derivative and chain rule on $F(\boldsymbol{u})$ with respect to u_i.

$$\frac{\partial F(\boldsymbol{u})}{\partial u_i} = \frac{\partial}{\partial u_i}\|\vec{I} - \vec{\eta}(\boldsymbol{u})\|^2 = -2\left(\frac{\partial}{\partial u_i}\vec{\eta}(\boldsymbol{u})\right)^T (\vec{I} - \vec{\eta}(\boldsymbol{u})). \tag{A.1}$$

The expression of $\frac{\partial}{\partial u_i}\vec{\eta}(\boldsymbol{u})$ can be simplified by noting that the dependence between u_i and both the expressions for $\vec{\eta}(\boldsymbol{u})$ defined in (18.a) and (18.b) becomes

$$\frac{\partial F(\boldsymbol{u})}{\partial u_i} = \begin{cases} S(\boldsymbol{u}_2)E^0\frac{\partial}{\partial u_i}\boldsymbol{u}_1 = S(\boldsymbol{u}_2)E^0(\vec{\epsilon}_{m\times 1})_i & 1 \le i \le m \\ E(\boldsymbol{u}_1)S^0\frac{\partial}{\partial u_i}\boldsymbol{u}_2 = E(\boldsymbol{u}_1)S^0(\vec{\epsilon}_{n\times 1})_{i-m} & m+1 \le i \le m+n \end{cases} \tag{A.2}$$

where $(\vec{\epsilon}_{m\times 1})_i$ and $(\vec{\epsilon}_{n\times 1})_j$ denote the $m \times 1$ and $n \times 1$ vectors with unity in their ith and j^{th} coordinates, respectively, and zeros elsewhere.

Let $M_{SE} = S(\boldsymbol{u}_2)E^0$ and $M_{ES} = E(\boldsymbol{u}_1)S^0$ be the $N \times m$ and $N \times n$ matrices, respectively. The more compact expression for $\frac{\partial\vec{\eta}(\boldsymbol{u})}{\partial u_i}$ is given by

$$\frac{\partial\vec{\eta}(\boldsymbol{u})}{\partial u_i} = \begin{cases} \vec{m}^i_{SE} & 1 \le i \le m \\ \vec{m}^{i-m}_{ES} & m+1 \le i \le m+n \end{cases} \tag{A.3}$$

where $\vec{m}^i_{SE} = M_{SE}(\vec{\epsilon}_{m\times 1})_i$ and $\vec{m}^j_{ES} = M_{ES}(\vec{\epsilon}_{n\times 1})_j$ are the ith and jth columns of M_{SE} and M_{ES}, respectively.

Substituting (A.3) into (A.1), the gradient $\nabla F(\boldsymbol{u})$ can then be expressed as

$$\nabla F(\boldsymbol{u}) = -2\begin{bmatrix} (\vec{m}^1_{SE})^T \\ \vdots \\ (\vec{m}^m_{SE})^T \\ (\vec{m}^1_{ES})^T \\ \vdots \\ (\vec{m}^n_{ES})^T \end{bmatrix}(\vec{I} - \vec{\eta}(\boldsymbol{u})) \tag{A.4.a}$$

$$= -2\begin{bmatrix} M^T_{SE} \\ \cdots \\ M^T_{ES} \end{bmatrix}(\vec{I} - \vec{\eta}(\boldsymbol{u})). \tag{A.4.b}$$

REFERENCES

[1] B. A. Wandell, "The synthesis and analysis of color image," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-9, pp. 2–13, 1987.

[2] L. T. Maloney, "Computational approaches to color constancy," Ph.D. Dissertation, Appl. Psychol. Lab., Stanford Univ., Stanford, CA, 1985.

[3] J. Ho, B. V. Funt, and M. S. Drew, "Separating a color signal into illuminant and surface reflectance components: Theory and applications," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 12, pp. 966–977, 1990.

[4] J. Ho, "Chromatic aberration: A new tool for color constancy," Master's Thesis, School Comput. Sci., Simon Fraser Univ., Vancouver, Canada, 1988.

[5] D. B. Judd, D. L. MacAdam, and G. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Amer. A*, vol. 54, pp. 1031–1040, 1964.

[6] J. Cohen, "Dependency of the spectral reflectance curves of munsell color chips," *Psychon. Sci.*, vol. 1, pp. 369–370, 1964.

[7] P.-R. Chang, T. H. Hsieh, and B. F. Yeh, "A color constancy model for advanced television cameras," *IEEE Trans. Broadcasting*, vol. 3, Jun. 1992.

[8] B. A. Wandell, "Color rendering of color camera Data," *Color Res. Appl.*, vol. 11, pp. S30-S33, 1986.

[9] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*. New York: Wiley, 1982, 2nd ed.

[10] G. Healey, "Color discrimination by computer," *IEEE Trans. Syst., Man Cybern.*, vol. 19, Nov. 1989.

[11] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverse and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, Apr. 1973.

[12] H. C. P. Berbee *et al.*, "Hit-and-run algorithm for the identification of nonredundant linear inequalities," *Math. Programming*, vol. 37, pp. 184–207, 1987.

[13] K. Ritter, "A method of conjugate directions for linearly constrained nonlinear programming problems," *SIAM Numer. Anal.*, vol. 12, no. 3, pp. 273–301, Jun. 1975.

[14] ——, "A superlinearly convergent method for minimization problems with linear inequalities constraints," *Math. Programming*, vol. 4, pp. 44–71, 1973.

[15] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1980, 2nd ed.

[16] P.-A. Wedin, "Perturbation bounds in connection with singular value decomposition," *BIT*, pp. 99–111, 1972.

[17] ——, "On pseudoinverse of perturbed matrix," Tech. Rep., Dept. of Comput. Sci., Lund Univ., Sweden, 1969.

[18] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Sci.*, vol. 220, pp. 671–680, 1983.

[19] F. Catthoor, H. deMan, and J. Vandewalle, "SAMURAI: A general and efficient simulated-annealing schedule with fully adaptive annealing parameters," *VLSI J.*, pp. 147–178, 1988.

[20] B. Hajek, "Cooling schedules for optimal annealing," *Math. Oper. Res.*, vol. 13, no. 2, pp. 311–329, May 1988.

[21] C. C. Skiscim and B. L. Golden, "Optimization by simulated annealing: A preliminary computational study for the TSP," in *Proc. 1983 Winter Simulation Conf.*, pp. 523–535.

[22] A. A. El Gamal, L. A. Hemachandra, I. Shperling, and V. K. Wei, "Using simulated annealing to design good codes," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 116–123, Jan. 1987.

[23] P. Carnevali, L. Coletti, and S. Patarnello, "Image processing by simulated annealing," *IBM J. Res. Develop.*, vol. 29, pp. 569–579, Nov. 1985.

[24] E. Aarts and J. Korst, *Simulated Annealing and Boltzman Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. New York: Wiley, 1989.

[25] S. Lang, *Linear Algebra*. Reading, MA: Addison-Wesley, 1972.

[26] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm," *ACM Trans. Math. Software*, vol. 13, pp. 262–280, Sept. 1987.

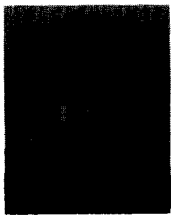
[27] H. Szu and R. Hartley, "Fast simulated annealing," *Phys. Lett.*, vol. 122, pp. 157–162, Jun. 1987.

[28] M. J. Vrhel and H. J. Trussell, "Color correction using principal components," *Color Res. Applicat.*, vol. 17, pp. 328–338, Oct. 1992.

[29] P.-R. Chang and T. H. Hsieh, "Constrained nonlinear optimization approaches to color-signal separation," Tech. Rep., TR/VSP-92-6, Dept. of Comm. Engr., Nat. Chiao-Tung Univ., Hsin-Chu, Taiwan, July 1992.

Po-Rong Chang (M'87) received the B.S. degree in electrical engineering from the National Tsing-Hua University, Taiwan, in 1980, the M.S. degree in telecommunications engineering from National Chiao-Tung University, Hsinchu, Taiwan, in 1982, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1988.

From 1982 to 1984, he was a lecturer in the Chinese Air Force Telecommunication and Electronics School for his two-year military service. From 1984 to 1985, he was an instructor of electrical engineering at National Taiwan Institute of Technology, Taipei, Taiwan. From 1989 to 1990, he was a project leader in charge of the SPARC chip design team at ERSO of the Industrial Technology and Research Institute, Chu-Tung, Taiwan. Currently, he is an Associate Professor of Communication Engineering at National Chiao-Tung University. His current interests include fuzzy neural networks, HDTV color signal processing, and human visual and audio systems.



Tsung-Hsieh Hsieh was born in Taiwan in 1968. He received the B.S. and the M.S. degrees in communication engineering from the National Chiao-Tung University, Hsin-chu, Taiwan, in 1990 and 1992, respectively.

His current research insterests are in the area of HDTV color signal processing, fuzzy set theory and applications, and neural networks.