# Bayes Information Criterion for Tikhonov Regularization with Linear Constraints: Application to Spectral Data Estimation

P. Carvalho, A. Dourado, B. Ribeiro*
CISUC - Centre for Informatics and Systems, University of Coimbra
Pólo II, Pinhal de Marrocos, 3030 Coimbra, Portugal
Email: {carvalho,dourado,bribeiro}@dei.uc.pt

## Abstract

*Spectral data estimation is an ill-posed problem, since (i) it is difficult to collect sufficient linear independent data and (ii) due to the integral nature of solid-state light sensors, camera outputs do not depend continuously on input signals. To solve these problems, most methods relay on exact a priori knowledge to reduce the problem's complexity (solution's space). In this paper a new algorithm is introduced which does not require a priori information. The method is build upon a new extension of the Bayes Information Criterion for ill-posed estimation problems, that is able to extract this information from the input data. Actually, the proposed solution is quite general and can readably be applied to other ill-posed problems, which are common in computer vision and image processing.*

## 1 Introduction

Many image processing and computer vision tasks require the estimation of spectral data. Usually, it involves the estimation, for each wavelength $\lambda$ (most often $\lambda$ is confined to the visible spectrum), of some data distribution $x(\lambda)$. For instance, many radiometric camera calibration [1], demosaicing, color constancy [2], and spectral reflectance estimation methods [3] require the knowledge of the spectral distribution of the light sensor's sensitivities. Further, as was pointed out by Hardeberg [4], sensor sensitivities are also important to improve color management systems for multimedia applications. Another typical situation where spectral data estimation has to be performed is reflectance estimation [3].

Spectral data estimation based on image data is an ill-posed problem, since (i) due to the integral nature of CCD and APS sensors, the same output can be obtained from an infinity of input signals, and (ii) it is observed that colors can be well approximated with just a few basis functions [3], which imposes an upper limit on the number of linear independent equations for the estimation problem. Fortunately, there are some assumptions that can be made on the solution to reduce the problem's complexity. The most commonly applied constraints are the solution's (i) positivity and (ii) smoothness [2][3][5], although other types of constraints can be found in literature: for instance, Finlayson [5] suggests using modality constraints, while local maxima are constrained in [6].There are several ways to impose the solution's smoothness. For instance, in [7], for each point, the absolute value of the second order derivative of the solution is required to be less than a predefined threshold, while Finalyson [5] uses projections onto a limited number of Fourier basis functions. If smoothness is encoded with a quadratic penalty involving derivatives such as $\alpha \int \left(x^{(k)}(\lambda)\right)^2 d\lambda$ ($x^{(k)}(\lambda)$ is the $k$th order derivative of the model and $\alpha$ is the regularization gain), then we have the so called Tikhonov regularization, a common technique to solve ill-posed problems.

Spectral data estimation can be formulated in therms of a least squares with linear inequality constraints problem [2][3][5][6]. In this paper the following formulation will be assumed: let $x \in \mathbb{R}^n$ be a discrete version of $x(\lambda)$ such that $x_i \equiv x(\lambda_i)$, $\lambda_i = \lambda_0 + (i-1)\Delta\lambda$, $i = 1...n$, and $\Delta\lambda$ is the sampling interval, then $x$ can be computed from

$$\min_x \left\{ \frac{\Delta\lambda^2}{m\sigma^2} \|Ax - y\|^2 + \frac{\alpha}{\Delta\lambda^4} \|Dx\|^2 \right\} \quad (1)$$

$$\text{subject to } \Xi \equiv \{C_i x \leq d_i, i = 1, ..., q\} \quad (2)$$

where $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$, $D \in \mathbb{R}^{l\times n}$, $C \in \mathbb{R}^{q\times n}$, $d \in \mathbb{R}^q$, $Ax = y$ are the $m$ equations (usually $m \ll n$) that can be obtained from the sensor's outputs, and $\alpha \in \mathbb{R}^+$ is a regularization gain that controls the trade-off between the roughness of the solution as measured by $\|Dx\|^2$ ($\Delta\lambda^{-2}Dx$ approximates the second derivative of $x$) and the infidelity to the data as measured by $\|Ax - y\|^2$. In (1) it is assumed that $y$ is subject to uncorrelated white

*Acknowledgements: To be included

noise, i.e., $y+\varepsilon{:}\varepsilon_i{\sim}\mathcal{N}(0,\sigma)$. If distinct noise variances are present in $y$ then (1) still holds if the following transformations are performed: let the noise associated to equation $i$, $i = 1...m$, be $\varepsilon_i{\sim}N(0,\sigma_i)$, then $A \equiv diag\,(\sigma/\sigma_i)\,A$ and $y \equiv diag\,(\sigma/\sigma_i)\,y$.

In practice, $C$, $d$ and $\alpha$ are defined based on exact *a priori* knowledge on $x$ [2][3][5] [6], being the solution very dependant on these values. In fig. 1 (b) sum squared errors ($SSE$) are depict as a function of the regularization gain for several CCD spectral sensitivity estimation problems. As can be observed, the quality of the solution is highly dependant on the quality of the *a priori* knowledge (in this case, the smoothness), which in most situations is difficult or even impossible to obtain. Hence, some validation tool is required to select these parameters. There are several measures based on the bias-variance/complexity trade-off principle (GCV [8], UBR [8], etc.), for regularization gain selection. However, these measures are not for problems with linear constraints. In [9] we extended GCV for this purpose. In this paper (section 2) a new selection method based on the Bayes Information Criteria, originally introduced by Schwartz and later extended by Neath and Cavanaugh [10] for model order selection in well-defined regression problems, is developed for model order and constraints parameterization selection. This criteria is applied in section 3 to spectral data estimation problems formulated in (1) and (2), namely to light sensor sensitivity estimation. Some main conclusions are also presented in section 3.

## 2 Extending the BIC for Model and Constraints Selection

The general Tikhonov regularization problem subject to linear inequality constraints in (1) and in (2), can be written as in 3.

$$\min_x \left\{ \|A(\gamma)\,x - \widetilde{y}\| \right\}, \text{ subject to } Cx \le d \qquad (3)$$

$$\gamma \equiv \frac{m\alpha\sigma^2}{\Delta\lambda^6}, A(\gamma) \equiv \begin{bmatrix} A \\ \sqrt{\gamma}D \end{bmatrix}, \widetilde{y} \equiv \begin{bmatrix} y \\ 0 \end{bmatrix}$$

Given a regularization gain $\gamma$ and a set of constraints $\Xi$, the optimal solution can be obtained using active set algorithms. However, in most situations, this *a priori* knowledge is unavailable. In this section the Bayes Information Criteria is extended in this context. Since the quadratic penalty in (1) can be seen as a quadratic constraint, for notation simplicity, all constraints are defined by $\Sigma \equiv \begin{pmatrix} \Xi & \gamma \end{pmatrix}$. Let $\theta = \begin{pmatrix} x^T & \sigma \end{pmatrix}$ be the vector which parametrizes each model, $Y$ be the set of data for the estimation problem, and $P(Y|\theta,\Sigma)$ the likelihood for $Y$ based on $\theta$ and $\Sigma$. Furthermore, let $P(\Sigma)$, $P(\theta|\Sigma)$ be prior distributions, respectively, for the constraints, and for the model's parameterization vector, given $\Sigma$. The joint a posteriori distribution $f(\theta,\Sigma|Y)$ for $\theta$ and $\Sigma$ given the observed data $Y$ can be obtained with the Bayes rule by (4) ($h(Y) \equiv \iint P(Y,\theta,\Sigma)\,d\theta d\Sigma$).

$$f(\theta,\Sigma|Y) = \frac{P(\Sigma)\,P(Y|\theta,\Sigma)\,P(\theta|\Sigma)}{h(Y)} \qquad (4)$$

Using this result, to select proper constraints $\Sigma$, one expects the a posteriori distribution $f(\Sigma|Y)$ to be maximized, i. e.,

$$\max_{\Sigma}\{f(\Sigma|Y)\} \Longleftrightarrow \min_{\Sigma}\left\{-2\ln\int f(\theta,\Sigma|Y)\,d\theta\right\} \quad (5)$$

In order to solve the integral in equation (5), the likelihood $P(Y|\theta,\Sigma)$ has to be determined. Let $\widehat{C}x = \widehat{d}$, $\widehat{C} \in \mathbb{R}^{p\times n}$, $rank(\widehat{C}) = p < n$ , be the active set of constraints at the solution of (3) computed for a particular set of parameters $\Sigma$. If these constraints are known, then the solution $\widehat{x}$ is easily computed as follows: let $A(\gamma) = U\begin{pmatrix} D_A^T & 0 \end{pmatrix}^T Z$ and $\widehat{C} = V\begin{pmatrix} D_C & 0 \end{pmatrix}Z$, be the generalized singular value decompositions of matrixes $A(\gamma)$ and $\widehat{C}$, where $D_A = diag\,(\alpha_1,...,\alpha_n)$, $D_C = diag\,(\beta_1,...,\beta_p)$, $U \in \mathbb{R}^{(m+l)\times(m+l)}$ and $V \in \mathbb{R}^{p\times p}$ are orthogonal matrixes and $Z \in \mathbb{R}^{n\times n}$ is invertible if $rank\left(A^T(\gamma) \;\; \widehat{C}^T\right)^T = n$. Using this matrix decomposition, it follows that $\widehat{x} = Z^{-1}\begin{pmatrix} y_1^T & y_2^T \end{pmatrix}^T, y_1 = D_C^{-1}V^T h \in \mathbb{R}^p$, and

$$\min_{y_2}\left\{\|Hy_2 - \widetilde{e}\|^2\right\} \qquad (6)$$

where, $D_{A_1} = diag\,(\alpha_1,...,\alpha_p)$, $D_{A_2} = diag\,(\alpha_{p+1},...,\alpha_n)$, $H = U\begin{pmatrix} 0 & D_{A_2}^T & 0 \end{pmatrix}^T \in \mathbb{R}^{(m+l)\times(n-p)}$ and $\widetilde{e} \equiv \widetilde{y} - U\left(\left(D_{A_1}D_C^{-1}V^T\widehat{d}\right)^T \;\; 0\right)^T$ (this result is partially proven in [9]).

Note that the constraints are subject to null noise, i. e., $\widehat{C}x = \widehat{h} + \epsilon$, $\epsilon \backsim \lim_{\sigma\to 0}\mathcal{N}\left(0,\sigma^2 I\right)$. Hence, $y_1 \backsim \lim_{\sigma\to 0}\mathcal{N}\left(D_C^{-1}V^T h,\Lambda\right) = \delta\left(y_1 - D_C^{-1}V^T h\right)$, where $\delta$ is the Dirac function and $\Lambda \equiv D_C^{-1}V\sigma^2 V^T D_C^{-1}$. Let $\theta^T = \begin{pmatrix} \theta_1^T & \theta_2^T \end{pmatrix}$, where $\theta_1 \equiv y_1$ is the section of the parameterization vector computed from the active constraints in $\Xi$. From (6) it is seen that $\theta_1$ and $\theta_2$ are independent. Hence, $P(\theta|\Sigma) = P(\theta_1|\Sigma)\,P(\theta_2|\Sigma) = \delta\left(\theta_1 - \widehat{\theta}_1\right)P(\theta_2|\Sigma)$. Substituting this result into (5) hilts

$$\begin{gathered} BIC_{IC}(\Sigma) = -2\ln P(\Sigma) + 2\ln h(Y) \\ -2\ln P\left(Y|\widehat{\theta}_1,\Sigma\right)\int P(Y|\theta_2,\Sigma)\,P(\theta_2|\Sigma)\,d\theta_2 \end{gathered} \qquad (7)$$

To compute the integral in (7), a second order Taylor approximation of $P(Y|\theta_2, \Sigma)$ can be used, as suggested in [10] for ordinary $BIC$. Let $\widehat{\theta}_2$ be the maximum likelihood estimate vector of $\theta_2$, then $\ln P(Y|\theta_2, \Sigma)$ can be approximated for a region near $\widehat{\theta}_2$ as in (8), where $F(\theta_2)$ is the Fisher information matrix defined in (9).

$$\begin{array}{c}\ln P(Y|\theta_2, \Sigma) \approx \ln P\left(Y|\widehat{\theta}_2, \Sigma\right) \\ -\frac{\left(\theta_2-\widehat{\theta}_2\right)^T F\left(\widehat{\theta}_2\right)\left(\theta_2-\widehat{\theta}_2\right)}{2(m+l)^{-1}}\end{array} \tag{8}$$

$$F(\theta_2) \equiv -\frac{1}{(m+l)} \frac{\partial^2 \ln P(Y|\theta_2, \Sigma)}{\partial\theta_2 \partial\theta_2^T} \tag{9}$$

Taking $P(\theta_2|\Sigma) = 1$, equation (10) follows (as observed in [10], this approximation holds as long as $P(Y|\theta_2, \Sigma)$ dominates the prior $P(\theta_2|\Sigma)$ within a small neighborhood of $\widehat{\theta}_2$; outside this neighborhood the exponential term in (8) should be small enough to force the product to zero).

$$\begin{array}{rl} & P\left(Y|\widehat{\theta}_2, \Sigma\right) \int P(Y|\theta_2, \Sigma) P(\theta_2|\Sigma)\, d\theta_2 \\ \approx & P\left(Y|\widehat{\theta}, \Sigma\right) (2\pi)^{\frac{n-p+1}{2}} \left|(m+l) F\left(\widehat{\theta}_2\right)\right|^{-\frac{1}{2}}\end{array} \tag{10}$$

Note that (6) implies that $\widetilde{e}_i = H_i y_2 + \epsilon$, $\epsilon \backsim \mathcal{N}\left(0, \sigma^2\right)$. Hence, the unbiased estimate of the variance that maximizes the likelihood is computed by (11).

$$\widehat{\sigma}^2 \equiv \frac{\left\|H\widehat{y}_2 - \widetilde{e}\right\|^2}{m+l} \tag{11}$$

On the other hand it is seen that

$$\left|F\left(\widehat{\theta}_2\right)\right| = 2(m+l)^{-n+p} \left(\widehat{\sigma}^2\right)^{-n+p-1} \left|H^T H\right| \tag{12}$$

Plugging (11), (10) and (12) into (7) and eliminating constant terms (13) is obtained.

$$\begin{array}{c} BIC_{IC}(\Sigma) = -2\ln P(\Sigma) + \ln\left|H^T H\right| \\ +(m+l-n+p-1)\ln\widehat{\sigma}^2 - (n-p)\ln(m+l) \\ +(n-p+1)\ln\frac{m+l}{2\pi}\end{array} \tag{13}$$

## 3 Results and Conclusions

In spectral data estimation, usually it is assumed that any radiometric distortions have been removed from the images. Hence, for a given input signal with power spectrum $e(\lambda)$ the pixel intensity can be assumed to be $\rho = \int e(\lambda) x(\lambda)\, d\lambda + \epsilon \approx \Delta\lambda \sum_{i=0}^{n} x_i e_i + \epsilon$ ($e_i \equiv e(\lambda_i)$, $\epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$). In the context of the sensor's sensitivity estimation, given a set of input signal $e_j(\lambda)$ and camera readings $\rho_j$, $j = 1, ..., m$, the $j$th line of matrix $A$ in (1) is defined by $A_j \equiv \begin{pmatrix} e_0 & ... & e_n \end{pmatrix}$ and $y_j \equiv \frac{\rho_j}{\Delta\lambda}$ (several variations of this formulation are possible, see for instance [2][9]). Photo-electrons of a given wavelength can be added, but never subtracted, which implies that $x_i \geq 0$, and, therefore, $C \equiv -I$ and $d \equiv 0$. As has been observed in [6], for small values of $\gamma$, the solution to this problem tends to exhibit large values of $SSE$ (Sum Squared Error) due to its oscillation. As $\gamma$ is increased, oscillation is eliminated. However, local maxima of the solution tend to be flatten due to the increased importance of the smoothness component in the objective function. This leads to increases of $SSE$ values. To measure the ability of the $BIC_{IC}$ criterion in identifying the optimal regularization gain $\gamma$, a simulation program was developed as described in [6]. In these simulations the prior $P(\Sigma)$ was modelled with $P(\Sigma) = P(\Xi) P(\gamma)$, where $P(\Xi) = 1$, $P(\gamma) = \mathcal{N}\left(\gamma_t, \sigma_t^2\right)$, $\sigma_t = \frac{\gamma_t}{3}$ and $\gamma_t$ is the lower limit of the identified stable regularization gain (gains which do not induce oscillation) as described in [6]. Further, the shown test results are for the spectral sensitivity curves from a Kodak DCS200 camera. In these testes $24$ ($m = 24$) patches of the MacBeth-Color Checker map were applied. Gaussian noise was added to each computed image. We modeled the noise's variance to be a linear function of $\rho$, such that $\sigma = 2$ for $\rho = 10$ and $\sigma = 6$ for $\rho = 250$. This is in accordance with real cameras [2]. Finally, the sampling step was fixed to $\Delta\lambda = 2nm$, $\lambda_0 = 400nm$, $\lambda_n = 700nm$ ($n = 151$) and the $SSE$ values were computed by $SSE = \left\|x_{True} - \widehat{x}\right\|^2$. A comparison between the described estimation method based on the $BIC_{IC}$ measure and by arbitrating $\gamma$ are depicted in fig. 1 for the RGB channels of the Kodak DCS200 camera. As can be observed (fig. 1 (a)-(b)) the $BIC_{IC}$ technique enables the estimation of suboptimal solutions in the vicinity of the global optimums. From fig. 1 (b) it is seen that for the red and for the green channels $SSE$ minimums occur at $\sqrt{\gamma} = 0.9$ and $\sqrt{\gamma} = 5.21$, while the selected $\gamma$ with the $BIC_{IC}$ criterion are $\sqrt{\gamma} = 0.81$ and $\sqrt{\gamma} = 4.81$ (fig. 1 (a)), respectively. This leads to similar $SSE$ values as the absolute minimums (fig. 1 (b)). As for the blue channel, the estimated $\gamma$ induces a larger $SSE$. However, as can be observed in fig. 1 (c), the larger $SSE$ is mainly due to the approximation error in the lower spectral range of wavelengths ($400nm$ to $420nm$), where light signal exhibit less energy, and, therefore, fewer information exists for the estimation problem. Actually, for larger values of $\gamma$ than $0.71$, $SSE$ decreases mainly due to a smoother approximation of the sensitivities in the above mentioned spectral region, and not due to a substantial improvement of the overall estimate.

In this paper a new extension to the Bayes Information Criterion is described for ill-posed problems (a relevant class of problems in computer vision and image processing tasks) that enables automatic selection of the model order and constraints parameterization. Based on this performance measure, a new spectral data estimation technique is introduced. No exact *a priori* knowledge on the data characteristics are required in this method (as opposed to other methods), since it is able to extract this information from the input data. This is a relevant result because, in practice, exact *a priori* knowledge is often difficult or even impossible to obtain with the required accuracy. The method is tested on light sensor spectral sensitivity estimation problems with unknown smoothness. The obtained results show that the $BIC_{IC}$ based method enables the identification of suboptimal solutions in the vicinity of the global optimum. More results will be shown in the final paper.

## References


[1] Y.-C. Chang and J. F. Reid, "RGB calibration for color image analysis in machine vision," *IEEE Trans. on Image Processing*, vol. 5, no. 19, pp. 1414–1422, 1996.

[2] K. Barnard, "Computational color constancy: Taking theory into practice," M.S. thesis, Simon Fraser University, 1995.

[3] J. Farrell, J. Cupitt, D. Saunders, and B. Wandell, "Estimating spectral reflectances of digital artwork," in *Chiba Confernce on Multispectral Imaging*, 1999.

[4] J. Hardeberg, *Acquisition and reproduction of colour images:colorimetric and multispectral approaches*, Ph.D. thesis, Ecole Nationale Superieure des Telecommunications, 1999.

[5] G. Finlayson and S. Hordley, "Recovering device sensitivity with quadratic programming," in *IST/SID's Color Imaging Conf.: Color Science, Systems and Applications*, 1998, pp. 90–95.

[6] Author, "To be included," in *To be included*, 2000.

[7] G. Sharma and H. Trussel, "Characterization of scanner sensitivity," in *Color Imaging Conf.: Transformations and Transportability of Color*, 1993, pp. 103–107.

[8] P. Craven and G. Wahba, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 31, pp. 377–403, 1979.

[9] Author, "To be included," in *To be included*, 2001.

[10] A. Neath and J. Cavanaugh, "Regression and time series model selection using schwartz information criterion," *Communications in Statistics Theory and Methods*, vol. 26, pp. 559–580, 1997.


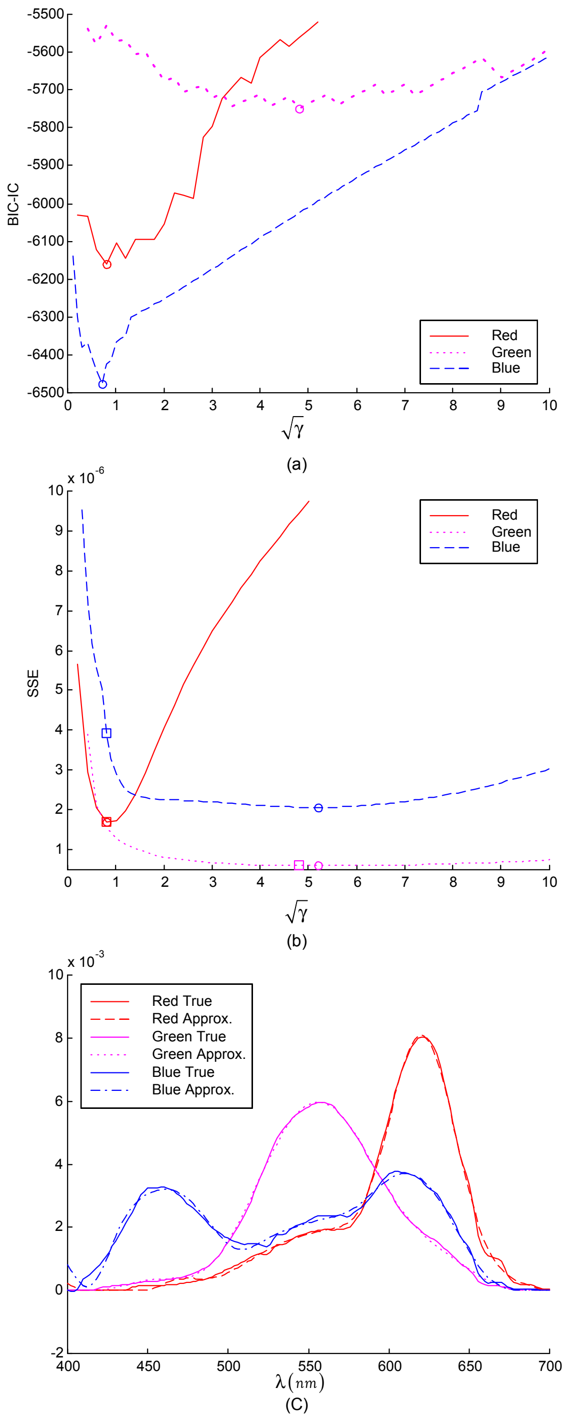


**Figure 1.** Kodak DCS200 sensitivity estimation results. (a) Evolution of $BIC_{IC}$ with $\gamma$. Minimums marked with circles. (b) SSE evolution with $\gamma$. Minimums marked with circles and selected $\gamma$ with $BIC_{IC}$ marked with squares. (c) Real and estimated sensitivities with $BIC_{IC}$.