光谱 Neugebauer 模型的抗差最小二乘优化

韩芳芳，李丽霞，张逸新
(江南大学，无锡 214122)

摘要： Yule-Nielsen 修正光谱 Neugebauer 模型是一种用来预测彩色网目调印刷品颜色的重要光谱预测模型。为了提高 Neugebauer 模型预测精度，在分析光谱 Neugebauer 模型和抗差估计理论的基础上，提出了将抗差最小二乘法用于计算模型参数网点面积率和基色光谱反射率。常用的抗差估计方案有：Huber 估计和 IGG 估计。实验表明，当存在粗差的情况下，与传统的最小二乘法相比，抗差最小二乘法的光谱 Neugebauer 模型的预测精度更高且更为稳定。

关键词： Neugebauer 模型；抗差最小二乘法；色彩管理
中图分类号： TS801.3；TS807 **文献标识码：** A **文章编号：** 1001-3563(2011)21-0106-04

Robustified Least Squares for Spectrum Neugebauer Model

HAN Fang-fang，*LI Li-xia*，*ZHANG Yi-xin*
(Jiangnan University，Wuxi 214122，China)

Abstract: Yule-Nielsen modified spectrum Neugebauer model is an important spectrum prediction model in predicting color output of color halftone prints. In order to improve the prediction accuracy，spectrum Neugebauer model and robustified estimation theory were analyzed and a method of robustified least squares to estimate dot area function and primary spectrum reflectance were put forward. Commonly used robust estimation programs are Huber estimation and IGG estimation. Experiment results showed that compared to traditional least squares (LS) based methods，the accuracy of spectrum Neugebauer model with RLS approach is higher and more stable under the condition of gross error.

Key words: Neugebauer model；robustified least square；color management

随着计算机和复制技术的发展，各种彩色图像的精确复现变得越来越重要。电子彩色设备大多采用与设备相关的色彩空间，如彩色扫描仪、显示器、打印机等，会导致彩色图像在各种设备之间的传输和复制在线过程中颜色失真，因而，色彩管理技术通过引入与设备无关的色彩空间，对设备进行特征化处理，以达到各种电子彩色设备之间的色彩的一致性。

色彩管理中印刷设备特征化的核心技术就是颜色空间转换。目前，颜色空间转换模型大致可分为理论模型法和经验模型法，前者需要充分的理论依据和数学公式来支持，但只需要较少量的测量数据；而后者不需要充分的理论依据，但需要测量大量的数据，通过各种数值算法拟合两颜色色空间的转换关系。文献[1－5]分别采用不同的方法修正了光谱 Neugebauer 模型，研究了彩色打印机的特征化问题。在研究中运用传统最小二乘法及整体最小二乘法对光谱 Neugebauer 模型的参数进行了优化处理，但是此 2 种方法没有考虑到测量光谱反射率相关的不确定信息，因此，得到的模型对由不可避免的空间的不均匀性及色彩偏差引起的光谱数据的误差是敏感的。

将抗差理论引入到光谱 Neugebauer 模型的参数估计当中，采用抗差最小二乘法计算模型参数，降低实验数据中粗差的影响，提高模型的预测精度和稳定性。

1 Yule-Nielsen 修正光谱 Neugebauer 模型

在考虑油墨层及纸张的光散射和吸收作用及其它工艺影响因素的情况下，半色调图像的光谱反射率

收稿日期： 2011-06-08
作者简介： 韩芳芳(1987－)，女，河南汝南人，江南大学硕士生，主攻印刷控制技术。

可用 Yule-Nielsen 修正光谱 Neugebauer 方程来描述[6]：

$$r(\lambda)^{\frac{1}{n}}=\sum_{p=1}^{8}a_p r_p(\lambda)^{\frac{1}{n}} \tag{1}$$

式中：$r(\lambda)$是面积元的光谱反射率；$r_p(\lambda)$是 Neugebauer 色元的光谱反射率；a_p 表示各 Neugebauer 色元在构成面积元时所占的比率；n 值被称为 Yule-nielsen 修正因子[6]。参数 n 是一个经验数值，取决于纸张的特性，加网线数等多方面因素。因此，可以把 n 值作为一个自由参数，使其扩展到一个很大的的区间，采用实验法或优化法确定最佳的 n 值，其修正效果最好。

2　Neugebauer 参数的抗差最小二乘优化

由于 Neugebauer 模型的参数是根据印刷品颜色的光谱测量值决定的，但光谱测量值本身含有误差，即当光谱测量值含有粗差(非正常值)时，采用传统最小二乘法计算，模型精度将会降低，因此采用抗差最小二乘法计算模型参数[2−5,7−8]，可以减免粗差的影响。

2.1　网点面积率

光散射效应以及油墨扩展导致了打印机的数字驱动值与模型中的网点面积率之间成非线性关系。既可以通过直接测量得到网点面积率，也可以把与输入数字驱动值相对应的网点面积率作为参数，通过拟合模型得到最优值，笔者采取后一种方法。

抗差最小二乘法方法可以用于计算单色印刷的网点面积率，以青色为例，青色的数字驱动值逐步从 0 到 255 的彩色梯尺区中，得到的单色印刷品，Neugebauer 基色只有 2 种：青色和白色。此时，Neugebauer 方程为：

$$r(\lambda)^{\frac{1}{n}}=(1-c)r_{\mathrm{w}}(\lambda)^{\frac{1}{n}}+cr_{\mathrm{c}}(\lambda)^{\frac{1}{n}} \tag{2}$$

经整理可得：

$$r(\lambda)^{\frac{1}{n}}-r_{\mathrm{w}}(\lambda)^{\frac{1}{n}}=c[r_{\mathrm{c}}(\lambda)^{\frac{1}{n}}-r_{\mathrm{w}}(\lambda)^{\frac{1}{n}}]$$

最终化为：

$$\hat{r}(\lambda)=c\hat{r}_{\mathrm{cw}}(\lambda)$$

式中：$\hat{r}(\lambda)=r(\lambda)^{\frac{1}{n}}-r_{\mathrm{w}}(\lambda)^{\frac{1}{n}}$；$\hat{r}_{\mathrm{cw}}(\lambda)=r_{\mathrm{c}}(\lambda)^{\frac{1}{n}}-r_{\mathrm{w}}(\lambda)^{\frac{1}{n}}$；$r(\lambda)$为印刷品的反射光谱率；$r_{\mathrm{w}}(\lambda)$为 Neugebauer 白基色的光谱反射率；$r_{\mathrm{c}}(\lambda)$为 Neugebauer 青基色的光谱反射率；$c$ 为青色的网点面积率。

对于三色印刷，以青色为例，青色的数字驱动值从 0～255 逐步变化，品红、黄两色的数字驱动值不变，以此得到印刷品，Neugebauer 基色有 8 种：青、品红、黄、红、绿、蓝、黑、白。此时，青色的网点面积率为所求参数，而品红和黄色的网点面积率为常量。Neugebauer 方程为：

$$r(\lambda)^{\frac{1}{n}}=c\sum_{k\in Sc}\hat{w}_k r_{pk}(\lambda)^{\frac{1}{n}}+(1-c)\sum_{l\in Snc}\hat{w}_l r_{pl}(\lambda)^{\frac{1}{n}} \tag{3}$$

式中：$\hat{w}_k=\dfrac{w_k}{c}$；$\hat{w}_l=\dfrac{w_l}{1-c}$。

在式(3)的右边，把含有 c 的因子与含有$(1-c)$的因子分别组合在一起。其中，c 为青色油墨的网点面积率；Sc 所对应的是含有青色的 Neugebauer 基色；Snc 所对应的是不含有青色的 Neugebauer 基色；$r(\lambda)$为印刷品的光谱反射率；$r_{pk}(\lambda)$与 $r_{pl}(\lambda)$为 Neugebauer 基色的光谱反射率；w_k 与 w_l 为 Neugebauer 基色的网点面积率。

经整理可得：

$$r(\lambda)^{\frac{1}{n}}-\sum_{l\in Snc}\hat{w}_l r_{pl}(\lambda)^{\frac{1}{n}}=c\Big[\sum_{k\in Sc}\hat{w}_k r_{pk}(\lambda)^{\frac{1}{n}}-\sum_{l\in Snc}\hat{w}_l r_{pl}(\lambda)^{\frac{1}{n}}\Big] \tag{4}$$

令 $\hat{r}(\lambda)=r(\lambda)^{\frac{1}{n}}-\sum_{l\in Snc}\hat{w}_l r_{pl}(\lambda)^{\frac{1}{n}}$；$\hat{r}_{pc}(\lambda)=\sum_{k\in Sc}\hat{w}_k r_{pk}(\lambda)^{\frac{1}{n}}-\sum_{l\in Snc}\hat{w}_l r_{pl}(\lambda)^{\frac{1}{n}}$。

最终化为：

$$\hat{r}(\lambda)=c\hat{r}_{pc}(\lambda) \tag{5}$$

设有一组相互独立的观测子样$\{\hat{r}_i\}$，$i=1,2\cdots n$，各观测子样的权为$\{p_i\}$，根据测量平差原理可得观测方程为：

$$\hat{\boldsymbol{R}}+\boldsymbol{\Delta}=\hat{\boldsymbol{R}}_{pc}c \tag{6}$$

式中：$\hat{\boldsymbol{R}}$ 为 $n\times 1$ 阶观测向量；$\hat{\boldsymbol{R}}_{pc}$ 为 $n\times 1$ 阶系数矩阵；c 为 1×1 阶未知参数网点面积率；$\boldsymbol{\Delta}$ 为 $n\times 1$ 阶观测误差向量。

因观测子样的真误差为未知，故相应的误差方程式可写为：

$$\boldsymbol{V}=\hat{\boldsymbol{R}}_{pc}c-\hat{\boldsymbol{R}} \tag{7}$$

式中：$\boldsymbol{V}$ 为观测值改正数(余差)向量。

由于抗差最小二乘估计属于广义极大似然(M 估计)估计范畴，且 M 估计通过观测值$\{\hat{r}_i\}$求参数 c 的估值，因此抗差最小二乘估计可由 M 估计准则函数来定义，即：

$$\sum_{i=1}^{n}p_i\left[\frac{\partial}{\partial v_i}\rho(v)\right]\frac{\partial v_i}{\partial c}=\sum_{i=1}^{n}p_i\phi_i(v)\hat{\boldsymbol{r}}_{pci}=0 \tag{8}$$

式中：$\rho(v)$ 是适当选择的凸函数；$\phi(v)=\frac{\partial}{\partial v_i}\rho(v)$ 是适当的单调、正半轴非降函数；v_i 是 i 个观测值余差；$\hat{\boldsymbol{r}}_{pci}$ 是系数阵 $\hat{\boldsymbol{R}}_{pc}$ 的第 i 行向量。

令 $\bar{p}_i=p_i\left(\frac{\partial p}{\partial v}\frac{1}{v}\right)_i=p_i\frac{\phi_i(v)}{v_i}=p_i w_i$，称 $\bar{p}_i$ 为等价权；w_i 为权因子，则：

$$\sum_{i=1}^{n} w_i v_i \hat{\boldsymbol{r}}_{pci}=0 \tag{9}$$

其矩阵表达式为：

$$\hat{\boldsymbol{R}}_{pc}{}^{\mathrm{T}}\bar{\boldsymbol{P}}\boldsymbol{V}=0 \tag{10}$$

由此得到抗差最小二乘估值：

$$c=(\hat{\boldsymbol{R}}_{pc}{}^{\mathrm{T}}\bar{\boldsymbol{P}}\hat{\boldsymbol{R}}_{pc})^{-1}\hat{\boldsymbol{R}}_{pc}{}^{\mathrm{T}}\bar{\boldsymbol{P}}\hat{\boldsymbol{R}} \tag{11}$$

由于观测值相互独立，故 $\bar{\boldsymbol{P}}$ 为对角阵，其对角元素为 $p_i w_i$。

由式(11)就可得到最优网点面积率 c，此种方法亦可适用于品红色和黄色，其每种颜色的最优网点面积率都可在确保其他颜色的网点面积率已知的情况下采用抗差最小二乘法估计得到。

抗差最小二乘估计是通过等价权将抗差估计与最小二乘法有机结合起来的结果，因此，抗差估计的根本之处在于权函数的设计。权函数的作用在于：将所有的观测数据划分为正常观测值、可利用观测值和粗差观测值 3 个部分；对于正常观测值，使其保持原有的权不变，对余差较大的可利用观测值进行降权处理，而那些粗差观测值，则使其权为 0。

常见的权函数为 Huber 估计、丹麦法、IGG 法、Hampel 三截尾估计、Andrews 正弦估计、Tukey 双权估计等多种方案。笔者仅用 Huber 估计和 IGG 算法进行对比计算。

$$\text{Huber 法},w(u)=\begin{cases}u & |u|\leqslant k\\ k\cdot\operatorname{sign}(u) & |u|>k\end{cases}$$

其中：$u=(\hat{r}-\hat{\boldsymbol{r}}_{pc}{}^{\mathrm{T}}c)/s$；$\hat{\boldsymbol{r}}_{pc}{}^{\mathrm{T}}$ 为系数阵 $\hat{\boldsymbol{R}}_{pc}$ 的行向量

s 在 $|u|\leqslant k$ 区间取 $\hat{\sigma}$(标准差)，$\hat{\sigma}=[(\hat{r}-\hat{\boldsymbol{r}}_{pc}{}^{\mathrm{T}}c)^2/(n-1)]^{\frac{1}{2}}$。

s 在 $|u|>k$ 取 MAD，$\mathrm{MAD}=\underset{i}{\mathrm{med}}|\hat{r}_i-\hat{\boldsymbol{r}}_{pc}{}^{\mathrm{T}}c|$，med 为中位数。

$$\text{IGG 法},w(u)=\begin{cases}1 & (|u|\leqslant 1.5)\\ k/|u| & (1.5<|u|\leqslant 2.5)\\ 0 & (|u|>2.5)\end{cases}$$

2.2 Neugebauer 基色光谱反射率

对于三色印刷的印刷色块，其 Neugebauer 方程为式(1)，且由式(1)可知 $r(\lambda)^{\frac{1}{n}}$ 为 $r_p(\lambda)^{\frac{1}{n}}$ 的线性函数，则式(1)可转化为此种形式 $R(\lambda)=\boldsymbol{A}\cdot R_p(\lambda)$，$R(\lambda)$为 Yule-nielsen 修正的印刷品的光谱反射率，$\boldsymbol{A}$ 为各 Neugebauer 基色的网点面积率函数矩；$R_p(\lambda)$为 Yule-nielsen 修正的各 Neugebauer 基色的光谱反射率。假设 $R(\lambda)$和 $\boldsymbol{A}$ 已知，$R(\lambda)$可直接测量得到，$\boldsymbol{A}$ 可通过采用抗差最小二乘法计算的网点面积率得到，由此就可通过抗差最小二乘法计算得到各 Neugebauer 基色的光谱反射率 $R_p(\lambda)$。

设有一组相互独立的观测子样$\{r_i\}$，$i=1,2\cdots n$，各观测子样的权为$\{p_i\}$，其相应的误差方程式为：

$$\boldsymbol{V}=\boldsymbol{A}\boldsymbol{R}_p-\boldsymbol{R}$$

式中：$\boldsymbol{R}$ 为 $n\times1$ 观测向量；$\boldsymbol{A}$ 为 $n\times8$ 系数矩阵；$\boldsymbol{R}_p$ 为 8×1 阶未知参数向量；$\boldsymbol{V}$ 为观测值改正数(余差)向量。其求解过程与网点面积率的求解过程类似，最终可得其抗差最小二乘解为：

$$\boldsymbol{R}_p=(\boldsymbol{A}^{\mathrm{T}}\bar{\boldsymbol{P}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\bar{\boldsymbol{P}}\boldsymbol{R}$$

$$\text{Huber 法},w(u)=\begin{cases}u & (|u|\leqslant k)\\ k\operatorname{sign}(u) & (|u|>k)\end{cases}$$

其中：$u=(r-\boldsymbol{a}^{\mathrm{T}}r_p)/s$，$\boldsymbol{a}^{\mathrm{T}}$ 为系数阵 $\boldsymbol{A}$ 的行向量，s 在 $|u|\leqslant k$ 区间取 $\hat{\sigma}$(标准差)，$\hat{\sigma}=[(r-\boldsymbol{a}^{\mathrm{T}}r_p)^2/(n-1)]^{\frac{1}{2}}$；$s$ 在 $|u|>k$ 取 MAD，$\mathrm{MAD}=\underset{i}{\mathrm{med}}|r_i-\boldsymbol{a}^{\mathrm{T}}r_{\mathrm{p}}|$，med 为中位数。

$$\text{IGG 法},w(u)=\begin{cases}1 & (|u|\leqslant 1.5)\\ k/|u| & (1.5<|u|\leqslant 2.5)\\ 0 & (|u|>2.5)\end{cases}$$

网点面积率的最初值是采用 Neugebauer 基色光谱反射率的直接测量值计算的，由于采用抗差最小二乘法计算的 Neugebauer 基色光谱反射率肯定与其直接测量值是不同的，因此可以采用抗差最小二乘法计算的 Neugebauer 基色光谱反射率重新计算网点面积率，重复进行迭代直到获得最优的网点面积率和 Neugebauer 基色光谱反射率，使得模型误差最小。

3 实验验证

由于模型参数的确定和模型精度的验证都需要大量的实验数据来进行，因此采用喷墨打印机印制 2 种样本色块。第 1 种是训练样本，共 102 个色块，用

于参数估计。训练样本色块包括 2 种梯尺：一种是单色梯尺，其数字驱动值从 0 到 255 取 17 个值；另一种是三色梯尺，其中一种颜色的数字驱动值对应于单色梯尺从 0～255 取 17 个值，另外 2 种颜色的数字驱动值为常量。第 2 种是测试样本，包含 125 个色块，其 c, m, y 的数字驱动值分别取自于{0，64，128，192，255}，用于验证模型精度。采用 X＝Rite SP62 测量样本色块在 400～700 nm 之间的光谱反射率，以 20 nm 为光谱间隔，取 16 个波段采样点的反射率。根据模型原理和计算所得参数值，通过测试样本来验证该模型。比较每个色块预测计算得出的光谱数据与实际测量得出的光谱数据，并求出 2 组数据的根均方差 RMSE。通过上述方法对模型的可行性与有效性作出评价，最后验证结果见表 1，可以看出采用传统最小二乘法(LS)计算，其光谱预测值容易受到误差的影响且其光谱预测精度低于采用抗差最小二乘法(RLS)计算所得到的；而采用抗差估计方法能取得比较令人满意的结果，并且其中采用 IGG 法比采用 Huber 法的光谱预测精度更为稳定，抗差性更好。

表 1　125 个色块的根均方误差的平均值、最大值和最小值

Tab. 1 Average, maximum, and minimum of RMSE of 125 color blocks

	LS 估计	Huber 估计	IGG 估计
平均值	0.049 7	0.025 6	0.019 8
最大值	0.098 8	0.031 9	0.024 0
最小值	0.022 2	0.018 4	0.015 3

4　结论

研究了采用抗差最小二乘法和光谱 Neugebauer 模型进行彩色打印机颜色特征化的问题。通过对比抗差最小二乘法与传统最小二乘法的计算结果，得到将抗差理论引入到光谱 Neugebauer 模型的参数估计当中，其光谱模型的预测精度要比采用传统最小二乘法的光谱模型的预测精度更高且更为稳定。

参考文献：

[1] BALASUBRAMANIAN R. Optimization of the Spectral Neugebauer Model for Printer Characterization[J]. J Electron Imaging，1999，8(2)：156－166.

[2] XIA Ming-hui，SABER E，SHARMA G，et al. End-to-end Color Printer Calibration by Total Least Squares Regression [J]. IEEE Trans Image Process，1999，8(5)：700－716.

[3] BALASUBRAMANIAN R. The Use of Spectral Regression in Modeling Halftone Color Printers[C]. in Proc IS&T/OSA Annual Conference，Optics&Imaging in the Information Age，Rochester，NY，1996：372－375.

[4] LANA C，ROTEA M，VIASSOLO D. Robust Estimation Algorithm for Spectral Neugebauer Models[J]. Journal of Electronic Imaging，2005，14(1)：1－10.

[5] URBAN P，GRIGAT Rolf-rainer. Spectral-based Color Separation Using Linear Regression Iteration[J]. Color Research and Application，2006，31：229－238.

[6] WYBLE D R，BERNS R S. A Critical Review of Spectral Models Applied to Binary Color Printing[J]. Color Research and Application，2000，25(1)：4－19.

[7] 胡成发. 印刷色彩与色度学[M]. 北京：印刷工业出版社，1993.

[8] 周江文，杨元喜，黄幼才. 抗差最小二乘法[M]. 武汉：华中理工大学出版社，1997.

[9] 杜艳君，张逸新. 经典色彩呈色模型[J]. 包装工程，2006，27(2)：99－101.

（上接第 105 页）

参考文献：

[1] 肖朝晖，纪钢. 条码技术及其在包装防伪中的应用[J]. 包装工程，2005，26(3)：68－70.

[2] 张玲，胡东红，孔华锋，等. 二维码码图结构特性分析[J]. 湖北大学学报(自然科学版)，2004，26(3)：226－231.

[3] 宋卫海，宋士银. 二维条码 PDF417 码技术的应用及发展前景[J]. 山东轻工学院院报，2001，15(3)：19－24.

[4] 周利红，刘书家. QR 码图像处理和译码方法研究[J]. 北京工商大学学报(自然科学版)，2008，26(1)：63－66.

[5] 梁治寇. 条码：防伪多面手[J]. 中国防伪报道，2009(2)：12－16.

[6] 冯峻峰，侯德鹏，刘承仆. 二维条码防伪查询中的应用[J]. 企业标准化，2006(6)：24－25.

[7] 何绘宇，何建新. 二维码在证书防伪中的应用研究[J]. 中国计量，2008(12)：43.

[8] 徐仁艳. 探讨银行凭证加载防伪二维码[J]. 金融会计，2010(7)：38－40.

[9] 薛蓬，王达娅. 二维条码在商品防伪上的应用[J]. 科技信息，2009(33)：427.

[10] 黄颖为，龚小超. 二维条码技术及其在防伪中的应用[J]. 中国品牌与防伪，2007(7)：60－63.