MULTIRESOLUTIONAL REGULARIZATION OF LOCAL LINEAR REGRESSION OVER ADAPTIVE NEIGHBORHOODS FOR COLOR MANAGEMENT

Nasiha Hrustemovic and Maya R. Gupta

University of Washington
Dept. of Electrical Engineering
Seattle, WA 98195

ABSTRACT

A multiresolutional regularization method is proposed for local linear regression that regularizes local mean squared-error by the mean squared-error of a larger neighborhood. The approach is similar in motivation to generalized Tikhonov regularization, but because the regularization trades-off between two like quantities, it is easier to interpret and specify the regularization parameter. Color management experiments with printers are used to compare the proposed regularized local linear regression to ridge regularization and generalized Tikhonov regularization. The local linear regressions use previously-validated adaptive neighborhoods. Results show that significant reductions in error can be achieved over the state-of-the-art.

Index Terms— color management, local linear regression, ridge regression, regularization, printers

1. INTRODUCTION

Quality color reproduction requires accurate device characterization. A standard approach to achieving this is to empirically characterize the way a device transforms input color signals to output colors, and model the inverse transformation by fitting a function to the sample color pairs. For printers this empirical characterization method has been shown to be particularly effective [1]. The set of training color sample pairs is obtained by printing RGB color patches and measuring printed colors in CIELab color space. The common data fitting algorithms used with the empirical approach are local linear regression, polynomial regression, neural networks, and spline fitting [1]. It was shown that among these methods, local linear regression has the best performance for a number of printers [1].

In previous work, we proposed using regularized local linear regression (specifically, ridge regression) over enclosing neighborhoods whose size adapts based on the local spatial distribution of training samples [2]. Experiments with three printers (two inkjets and one xerographic) showed that using enclosing neighborhoods is an effective alternative to using a fixed neighborhood size, that the ridge regression yields improvements over linear regression, and that the lowest error was achieved using both an enclosing neighborhood and ridge regression.

By definition, ridge regression shrinks each component of the linear regression coefficient vector towards zero. The amount of shrinkage towards zero is determined by the regularization parameter λ. In this paper, we propose a new form of regularized linear regression, *local-global regression*, such that the regularization parameter regularizes the linear regression coefficients by taking into account the mean squared error for the entire training set. We also consider using a generalized Tikhonov regularization which shrinks the regression coefficients towards the regression coefficients which are the best fitted linear model for the whole training set. Experiments show that the multiresolutional regularization and generalized Tikhonov regularization can significantly decrease error compared to ridge regression.

In theory, an effective regularization parameter λ can be determined by cross-validation on a separate dataset. In many settings, cross-validation is impractical for device characterization, because of the large number of measurements required. In our previous work we selected the value $\lambda = .1$ based on the RGB errors from a small preliminary experiment. In this paper, we instead choose the regularization parameter based on the visual quality of four images.

Section 2 details the new and standard regularized local linear regression methods evaluated in our experiments. Section 3 reviews the adaptive neighborhood used to define locality. The experimental setup is explained in Section 4, and the results are provided in Section 5.

2. REGULARIZED LOCAL LINEAR REGRESSION

Local linear regression models a function by piecewise hyperplanes, given a set of n training sample pairs $\{(x_i, y_i)\}$ for $i = 1, \ldots, n$, where each input color $x_i \in \mathbb{R}^d$ and each corresponding output color $y_i \in \mathbb{R}^m$. To estimate the output value at point $g \in \mathbb{R}^d$, a set of training samples close to g is cho-

This research was supported by Sharp Labs of America.

sen from the n training samples; let k denote the number of neighbors in the neighborhood of g. Denote the jth neighbor v_j, and the jth neighbor's output color z_j. Then a hyperplane is fit to the k neighborhood training sample pairs $\{(v_j, z_j)\}$, and the hyperplane is evaluated at g to form the estimate $\hat{f}(g)$. In Section 3, we discuss the choice of the neighborhood set of training samples over which to fit the local hyperplane.

Local linear regression can have high estimation variance, and high estimation variance is often effectively controlled by regularization [3,4]. In previous work, we proposed regularizing local linear regression for color management; specifically, we showed that experimental error is lowered with ridge regression, a form of regularization that reduces the variance by penalizing the squared magnitude of the regression coefficients [2, 4]. Let $\bar{v}$ denote the $d \times 1$ mean vector of $\{v_j\}$, σ denote the $d \times 1$ standard deviation vector of $\{v_j\}$, and $\bar{z}$ denote the scalar mean of $\{z_j\}$. Let $\tilde{v}_j = (v_j - \bar{v})./\sigma$, $\tilde{g} = (g - \bar{v})./\sigma$, where $./$ denotes component-wise division. The component-wise scaling by the σ is a common approach to balance the effect of the regularization across the different input components [4]. Lastly, let $\breve{z}_j = z_j - \bar{z}$. Then local linear regression with ridge regularization estimates the output for g as $\hat{f}(g) = \tilde{g}^T \beta_r + \bar{z}$, where

$$\beta_r = \arg\min_{\beta \in \mathbb{R}^d} \sum_{j=1}^{k} \left(\tilde{v}_j^T \beta - \breve{z}_j\right)^2 + \lambda \beta^T \beta, \qquad (1)$$

where $\lambda \in \mathbb{R}^+$ is the regularization parameter.

Ridge regression pushes the regression coefficients towards zero, making the fitted model flatter than it would be without regularization. Larger values of λ result in lower estimation variance but higher estimation bias. Color transformations are usually monotonically increasing, and thus we hypothesize that making the model flatter will reduce variance at the cost of significant bias. Instead, we propose to regularize towards the global trend of the device's color transformation. One method to do this is generalized Tikhonov regularization, which regularizes towards a fixed regression vector, rather than towards zero. Let β_n be the least-squares linear regression coefficient vector fitted to all n training sample pairs $\{(x_i, y_i)\}$ for $i = 1, \ldots, n$. Then the local linear regression with Tikhonov regularization towards β_n is $\hat{f}(g) = \tilde{g}^T \beta_t + \bar{z}$, where

$$\beta_t = \arg\min_{\beta \in \mathbb{R}^d} \sum_{j=1}^{k} \left(\tilde{v}_j^T \beta - \breve{z}_j\right)^2 + \lambda \left(\beta - \beta_n\right)^T \left(\beta - \beta_n\right).$$

The closed form solution for β_t is given by,

$$\beta_t = \left(\tilde{V}^T \tilde{V} + \lambda I\right)^{-1} \left(\tilde{V}^T \breve{z} + \lambda \beta_n\right),$$

where $\tilde{V}$ is the $k \times d$ matrix with jth row $\tilde{v}_j^T$.

The form of the generalized Tikhonov regularization and ridge regularization make it difficult to choose a value a priori for the regularization parameter λ because λ trades-off between two very different quantities: total squared error of the fit, and squared error of the regression coefficients. To increase the interpretability of the regularization and thus make it easier to choose λ, we propose a multiresolutional alternate to generalized Tikhonov regularization that we term *local-global regression*. In local-global regression the parameter $\lambda \in [0, 1]$ controls the convex combination of two like quantities: local average-error and global average-error. Because the regularization is also in terms of output errors, there is no need to whiten. The proposed local-global linear model is:

$$\hat{f}(g) = g^T \beta_{LG} + (1 - \lambda)(\bar{z} - \bar{v}^T \beta_{LG}) + \lambda(\bar{y} - \bar{x}^T \beta_{LG}),$$

where the regression coefficients solve

$$\beta_{\text{LG}} = \arg\min_{\beta \in \mathbb{R}^d} \quad (1 - \lambda)\left(\tfrac{1}{k} \textstyle\sum_{j=1}^{k} \left(\breve{v}_j^T \beta - \breve{z}_j\right)^2\right) + \lambda\left(\tfrac{1}{n} \textstyle\sum_{i=1}^{n} \left(\breve{x}_i^T \beta - \breve{y}_i\right)^2\right)$$

for $\breve{v}_j = v_j - \bar{v}$; $\breve{x}_i = x_i - \bar{x}$ where $\bar{x}$ is the $d \times 1$ mean of $\{x_i\}$ for $i = 1, \ldots, n$; and $\breve{y}_i = y_i - \bar{y}$ where $\bar{y}$ is the scalar mean of $\{y_i\}$ for $i = 1, \ldots, n$.

The corresponding closed form solution is,

$$\beta_{\text{LG}} = \left((1 - \lambda)\tfrac{V^T V}{k} + \lambda \tfrac{X^T X}{n}\right)^{-1} \left((1 - \lambda)\tfrac{V^T z}{k} + \lambda \tfrac{X^T y}{n}\right),$$

where V is the $k \times d$ matrix with jth row $\breve{v}_j$ and z is the $k \times 1$ vector with jth component z_j; and X is the $n \times d$ matrix with ith row $\breve{x}_i$ and y is the $n \times 1$ vector with ith component y_i.

With ridge and Tikhonov regularization it is difficult to interpret the effect of a particular value of λ, as λ trades between two different types of quantities. This makes it difficult to select a λ or even to specify reasonable parameter choices to use in cross-validation for ridge or Tikhonov. In contrast, from (2), one sees that the local-global regularization parameter λ determines the percentage of local versus global error minimized, and as such, we hypothesize that this is an easier form of regularization to use in practice.

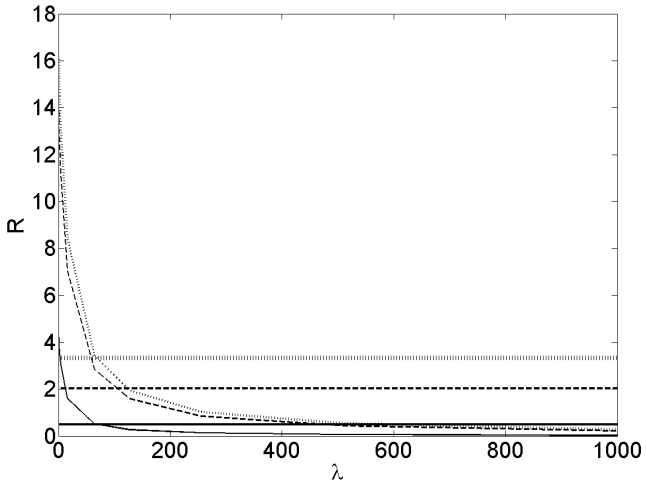
Examples of the different regression coefficients fitted for estimating one output color channel (R) are shown in Fig. 1 for ridge and local-global as the regularization parameter λ is increased. For our experiments, there are three input channels, L^*, a^* and b^*. In Fig. 1, β_{L^*} is represented with the dotted line, β_{a^*} is represented with the dashed line, and β_{b^*} is represented with the solid line.

3. ADAPTIVE NEIGHBORHOODS

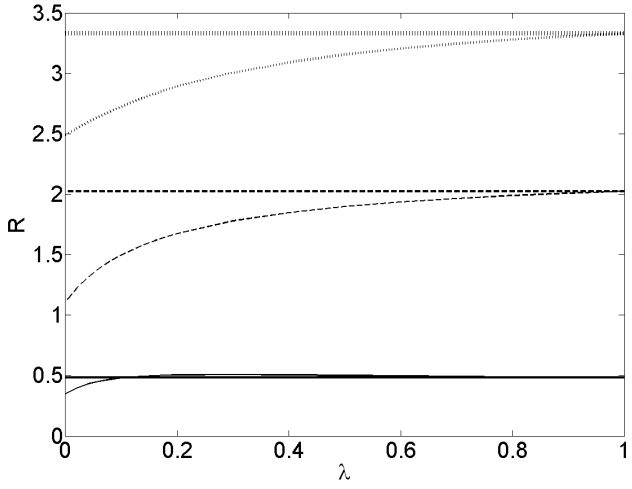
The set of neighbor training samples of a grid point g used for the local linear regression is commonly determined by using a fixed number k, chosen based on past experience or

cross-validation [4]. Cross-validation is costly for the device characterization problem because errors must be measured. Instead, as in our previous work, we use enclosing neighborhoods that automatically adapt their size based on the geometry of data: k-NN enclosing neighborhood.

The enclosing k-NN neighborhood is the smallest set of k-nearest neighbors that produces the minimal distance to enclosure, where the distance to enclosure is defined to be the Euclidean distance between g and the convex hull of the training samples in the neighborhood [2]. That is, if g is contained in the convex hull of all n training samples, then the enclosing k-NN neighborhood is the first k nearest neighbors whose convex hull contains g. An example of enclosing k-NN neighborhood is presented in Fig. 2. The figure shows the test point g and its neighboring points with the Voronoi tessellation diagram as a reference. More details about enclosing neighborhoods can be found in [2].



Ridge Coefficients



Local-Global Coefficients

Fig. 1. Profiles of the regularization coefficients as the regularization parameter λ is increased. The corresponding thicker lines show the values of the respective coefficients of the global least-squares solution β_n. The regression coefficient β_{L^*} is represented with the dotted line, β_{a^*} is represented with the dashed line, and β_{b^*} is represented with the solid line.

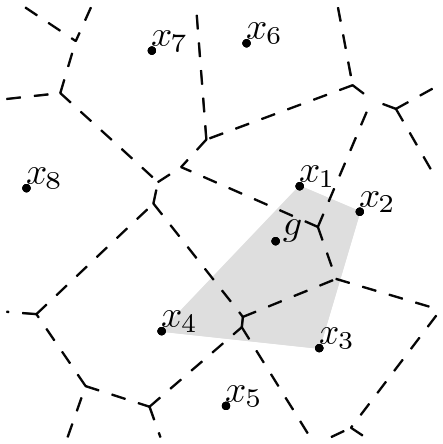


Fig. 2. Two-dimensional example of the *enclosing k-NN* neighbors of g, $\{x_i\}_{i=1,2,3,4}$.

4. COLOR MANAGEMENT EXPERIMENT

Each pairing of regularization method and neighborhood was tested on an HP Photosmart 720 ink jet printer and Samsung CLP-300 laser printer. The training color sample pairs were obtained by printing the 918 sample RGB Chromix test chart to produce $\{y_i\}$ for $i = 1, \ldots, 918$ with color components R, G and B. These 918 RGB values include 729 uniformly-spaced RGB colors that span the 8-bit RGB cube, and 189 additional neutrals and highly saturated primaries. Then the printed colors were measured to get the corresponding CIELab values $\{x_i\}$, with color components L^*, a^* and b^*. Color measurements of the printed samples were done with a GretagMacbeth Spectrolino spectrophotometer at a 2° observer angle with D50 illumination.

The color management system in our experiments used a standard empirical characterization implementation with a 3D look-up-table (LUT) that characterizes the color transformation from the desired CIELab color values to their corresponding RGB color values, and three 1D LUTs that prelinearize each channel independently. The 1D LUTs perform gray-balance calibration by enforcing that neutral RGB color values (R=G=B) produce gray colors such that the printed patches have measured CIELab values with $a^* = 0$ and $b^* = 0$ and uniform steps of L^* [1]. The domain of the 3D LUT is a $17 \times 17 \times 17$ set of points uniformly spaced in CIELab

color space such that $L^* \in [0, 100]$ and $a^*, b^* \in [-100, 100]$, and the LUT is centered on the neutral axis ($a^* = b^* = 0$).

LUTs were estimated for five regression methods. Each regression technique requires a regularization parameter λ. In our previous work we selected the value $\lambda = .1$ based on the RGB errors from a small preliminary experiment. In this paper, we instead chose the regularization parameter for each regularization method based on the visual quality of a set of four training images. We first generated LUTs using the sets of regularization parameters: $\lambda \in \{0.1, 1, 5, 10\}$ for ridge and Tikhonov, and $\lambda \in \{0.005, 0.01, 0.05, 0.1\}$ for local-global. The final regularization parameter for each method was selected based on visual inspection of 4 images from the public Kodak photo set: kodim04, kodim17, kodim20, and kodim23. Each image, initially specified in RGB, was converted to a CIELab image by assuming a D50 white point and using the standard sRGB-to-CIELab transform. Next, we used the generated LUTs to obtain the corresponding printer-RGB test images, printed the images, and performed a visual inspection for false contours. For each neighborhood-regression pair's LUT, we selected the smallest parameter λ out of the above set of possible values that produced no images with false contours.

Each test color was drawn randomly and uniformly from the RGB color cube, printed for each printer, and measured to determine the corresponding CIELab color. The 918 measured CIELab colors for each printer formed that printer's test sample set. This process guaranteed that the test samples were all in the gamut of the printer. The test samples were then input into the color management system with each neighborhood-regression pair's LUTs. The LUTs produced estimated RGB values that were then printed, and measured. The measured CIELab values were then compared to the test CIELab values, using ΔE^*_{94} error.

5. RESULTS

Table 1 and Table 2 show the average ΔE^*_{94} error, the 95^{th} percentile error, and the maximum error for each regression method for the HP inkjet and Samsung laser printer. For color management, a single poorly reproduced color is very noticeable, whereas a number of small errors may not be noticeable. Therefore we emphasize the 95^{th} %-ile error and maximum error as the more practical metrics.

As hypothesized, the Tikhonov and local-global regularizations provide consistent error reductions over ridge regression for both printers, validating the proposed use of the whole training set in the regularization. In particular, the local-global regression method shows the best performance on the mean and 95^{th} percentile error. Specifically, the local-global yields 17% lower average error, 24% lower 95^{th} %-ile error, and 35% lower maximum error compared to ridge on the HP inkjet. On the Samsung laser printer, all of the methods perform similarly, but the local-global still produces a 7.5% reduction in average error and a 10% reduction in 95^{th} %-ile error compared to ridge regression.

Table 1. ΔE^*_{94} errors for HP Photosmart D7260 inkjet printer

Regression Method		Average Error	95^{th} %-ile Error	Max Error
Regularization	Parameter			
Ridge	$\lambda = 1$	2.3	5.0	10.2
Local-global	$\lambda = .005$	1.9	3.8	6.6
Tikhonov	$\lambda = 1$	2.0	4.0	8.1

Table 2. ΔE^*_{94} errors for Samsung CLP-300 laser printer

Regression Method		Average Error	95^{th} %-ile Error	Max Error
Regularization	Parameter			
Ridge	$\lambda = 1$	4.0	8.4	15.6
Local-global	$\lambda = .01$	3.7	7.6	15.6
Tikhonov	$\lambda = 10$	3.7	7.9	15.6

6. REFERENCES

[1] R. Bala, *Digital Color Handbook*, chapter 5: Device Characterization, pp. 269–384, CRC Press, 2003.

[2] M. R. Gupta, E. K. Garcia, and E. M. Chin, "Adaptive local linear regression with application to printer color management," *IEEE Trans. on Image Processing*, vol. 17, no. 6, pp. 936–945, 2008.

[3] A. E. Hoerl and R. Kennard, "Ridge regression: biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.

[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.