

# Piecewise Multi-linear PDF Modelling, Using an ML Approach

Edgard Nyssen, Naren Naik, and Bart Truyen

Vrije Universiteit Brussel, Vakgroep Elektronica en Informatieverwerking (ETRO)  
Pleinlaan 2, B-1050 Brussel, Belgium  
`ehnyssen@etro.vub.ac.be`

**Abstract.** This paper addresses the problem of estimating the model parameters of a piecewise multi-linear (PML) approximation to a probability density function (PDF). In an earlier paper, we already introduced the PML model and discussed its use for the purpose of designing Bayesian pattern classifiers. The estimation of the unknown model parameters was based on a least squares minimisation of the difference between the estimated PDF and the estimating PML function. Here, we show how a Maximum Likelihood (ML) approach can be used to estimate the unknown parameters and discuss the advantages of this approach. Subsequently, we briefly introduce its application in a new approach to histogram matching in digital subtraction radiography.

## 1 Introduction

In an earlier paper [1], we already addressed the problem of estimating the class-conditioned probability density function (PDF)  $f(\bar{x}|\omega \in \Omega_t)$ , appearing in the expression of a Bayesian discriminant function  $d_t(\bar{x}) = P(\omega \in \Omega_t)f(\bar{x}|\omega \in \Omega_t)$ . We cited different approaches [2,3,4,5,6] to the solution of this problem and proposed an alternative representation of approximated PDFs  $f_t(\bar{x})$  defined in a bounded domain  $I$ . In this approach, the domain is divided into cells on the basis of a multidimensional rectangular point lattice. The probability densities inside the cells are obtained by a multi-linear interpolation of function values at the lattice points (i.e. inside a cell, and along any line segment parallel to one of the main axes of the coordinate system, values are obtained by linear interpolation). In [1], we showed that in a low-dimensional feature space, this interpolation model allows a fast approximation of a PDF value in any point of  $I$ , and unlike other models, the speed of the calculations is independent of the model complexity. The piecewise multi-linear (PML) function  $f_t(\bar{x})$ , which maps the points of the domain  $I$  to the interpolated values, and which serves as an approximation of  $f(\bar{x}|\omega \in \Omega_t)$ , is reformulated as a weighted sum of PML basis functions. This allows the application of a procedure to optimise the approximation. In [1], we considered the minimisation of the least squares (LS) fitting criterion

$$C = \int_{\bar{x} \in I} (f(\bar{x}|\omega \in \Omega_t) - f_t(\bar{x}))^2 d\bar{x} . \quad (1)$$

In the present paper, we will prove that the approximating function  $f_t(\bar{x})$ , obtained in this way, satisfies

$$\int_{\bar{x} \in I} f_t(\bar{x}) d\bar{x} = 1 . \quad (2)$$

We will also prove that this property holds both for the “theoretical” approximating function  $f_t(\bar{x})$  — which paradoxically requires the knowledge of the exact PDF  $f(\bar{x}|\omega \in \Omega_t)$  — and its estimation  $\hat{f}_t(\bar{x})$ , derived from the data in a learning set of patterns vectors. Although (2) suggests that the approximation has the same properties as a probability density function, another fundamental property of PDF’s, namely positivity, unfortunately is not always satisfied, as is illustrated by some of the examples given in [1]. If this positivity is crucial, one may consider to search for a solution by finding a minimum of  $C$  in (1) under the constraint of positivity. Unfortunately, this solution not only is suboptimal but also, may no longer satisfy (2).

An alternative is to use another meaningful optimisation criterion. In the present paper, we suggest a Maximum Likelihood approach (see e.g. [2,7]). We show how this criterion can be applied, by reformulating the basic problem in an appropriate way, and demonstrate that the corresponding solution satisfies all properties of a PDF. We then present very briefly results obtained by a new approach to the problem of histogram matching in digital subtraction radiography, based upon the piecewise linear (PL) approximation of the PDFs underlying the histograms.

## 2 Fundamental Considerations

Preliminary remark: since the present paper does not relate data with pattern classes, we will omit the class index  $t$ , in our notations. The approximated PDF will be denoted by  $f(\bar{x})$ , its approximation will be referred to as  $f'(\bar{x})$ , and the estimator of this approximation will be indicated by  $\hat{f}'(\bar{x})$ .

### 2.1 Fundamental Properties of the Approximating Functions Obtained by Applying the Least Squares (LS) Fitting Criterion

The properties of the approximating functions, proven here, are based on the existence of a decomposition of any given constant function into a given set of basis functions  $\psi_j$ ,  $j \in \{1, \dots, m\}$ . Obviously, this is the case with any PML model.

We will prove two fundamental properties regarding the approximation of a probability function  $f(\bar{x})$ , defined in a finite domain  $I$ , using weighted sums of basis functions. This approximation is given by

$$f'(\bar{x}) = \sum_{j=1}^m \alpha_j \psi_j(\bar{x}) , \quad (3)$$

where the weight coefficients  $\alpha_j$  minimise the criterion

$$C = \int_{\bar{x} \in I} (f(\bar{x}) - f'(\bar{x}))^2 d\bar{x} . \quad (4)$$

**Theorem 1.** *If there exists a set of coefficients  $b_i$ ,  $i \in \{1, \dots, m\}$ , satisfying  $\sum_{i=1}^m b_i \psi_i(\bar{x}) = 1$ ,  $\forall \bar{x} \in I$ , then the approximation  $f'(\bar{x})$  satisfies  $\int_{\bar{x} \in I} f'(\bar{x}) d\bar{x} = 1$ .*

*Proof.* Substituting  $f'(\bar{x})$  in (4), by the right hand side of (3), and equating to zero the derivative of  $C$ , with respect to  $\alpha_j$ , one obtains

$$\sum_{j=1}^m \alpha_j \int_{\bar{x} \in I} \psi_j(\bar{x}) \psi_i(\bar{x}) d\bar{x} = \int_{\bar{x} \in I} \psi_i(\bar{x}) f(\bar{x}) d\bar{x} , \quad \forall i \in \{1, \dots, m\} . \quad (5)$$

Now, multiplying both sides of this equation by  $b_i$ , calculating the sum over  $i \in \{1, \dots, m\}$  for both expressions, and rearranging the order of summation and integration operations yields

$$\int_{\bar{x} \in I} \sum_{j=1}^m \alpha_j \psi_j(\bar{x}) \sum_{i=1}^m b_i \psi_i(\bar{x}) d\bar{x} = \int_{\bar{x} \in I} \left( \sum_{i=1}^m b_i \psi_i(\bar{x}) \right) f(\bar{x}) d\bar{x} , \quad \forall i \in \{1, \dots, m\} .$$

Substituting the first sum by the left hand side of (3), this equation simplifies to

$$\int_{\bar{x} \in I} f'(\bar{x}) d\bar{x} = \int_{\bar{x} \in I} f(\bar{x}) d\bar{x} = 1 ,$$

since  $b_i$ ,  $i \in \{1, \dots, m\}$ , satisfy  $\sum_{i=1}^m b_i \psi_i(\bar{x}) = 1$ ,  $\forall \bar{x} \in I$ .  $\square$

When a representative learning sample of pattern vectors  $\bar{x}_l$  is available, the coefficients  $\alpha_j$  can be estimated by replacing the expression at the right hand side of (5) — which represents the expectation of the value of the basis function  $\psi_i(\bar{x})$  — by the sample mean value of this function, which gives

$$\sum_{j=1}^m a_j \int_{\bar{x} \in I} \psi_j(\bar{x}) \psi_i(\bar{x}) d\bar{x} = \frac{1}{p} \sum_{l=1}^p \psi_i(\bar{x}_l) , \quad \forall i \in \{1, \dots, m\} . \quad (6)$$

Here,  $p$  is the sample size and  $a_j$  are the estimations of the original coefficients  $\alpha_j$  in (3). When the coefficients  $\alpha_j$  in the decomposition of the PDF are substituted by the coefficients  $a_j$ , the resulting function becomes an estimator for  $f'(\bar{x})$ :

$$\hat{f}'(\bar{x}) = \sum_{j=1}^m a_j \psi_j(\bar{x}) . \quad (7)$$

A second theorem shows that the property, proven previously for  $f'(\bar{x})$ , also holds for  $\hat{f}'(\bar{x})$ :

**Theorem 2.** *If there exists a set of coefficients  $b_i$ ,  $i \in \{1, \dots, m\}$ , satisfying  $\sum_{i=1}^m b_i \psi_i(\bar{x}) = 1$ ,  $\forall \bar{x} \in I$ , then the estimated approximation  $\hat{f}'(\bar{x})$  satisfies  $\int_{\bar{x} \in I} \hat{f}'(\bar{x}) d\bar{x} = 1$ .*

*Proof.* Summation over  $i \in \{1, \dots, m\}$  of both sides of (6), after multiplication with  $b_i$ , and rearranging the order of summation and integration operations, yields

$$\int_{\bar{x} \in I} \sum_{j=1}^m a_j \psi_j(\bar{x}) \sum_{i=1}^m b_i \psi_i(\bar{x}) d\bar{x} = \frac{1}{p} \sum_{l=1}^p \left( \sum_{i=1}^m b_i \psi_i(\bar{x}_l) \right), \quad \forall i \in \{1, \dots, m\} .$$

After the substitution of the first sum by the left hand side of (7), this equation yields:

$$\int_{\bar{x} \in I} \hat{f}'(\bar{x}) d\bar{x} = \frac{1}{p} \sum_{l=1}^p 1 = 1 ,$$

since  $b_i$  satisfy  $\sum_{i=1}^m b_i \psi_i(\bar{x}) = 1$ ,  $\forall \bar{x} \in I$ , including the learning pattern vectors  $\bar{x}_l$ .  $\square$

## 2.2 Derivation of a Maximum Likelihood Model for Estimating the Coefficients of the PML Approximation

Let us assume that the basis functions  $\psi_j$  satisfy

$$\forall \bar{x} \in I : \psi_j(\bar{x}) \geq 0 \quad \text{and} \quad \int_{\bar{x} \in I} \psi_j(\bar{x}) d\bar{x} = 1 . \quad (8)$$

In other words, the  $\psi_j$  behave like probability density functions. If the second condition does not hold for basis functions  $\psi'_j$  in the decomposition

$$f(\bar{x}) = \sum_{j=1}^m \alpha'_j \psi'_j(\bar{x}) ,$$

it is sufficient to multiply these with an appropriate scale factor  $s$ , i.e.  $\psi_j(\bar{x}) = s \psi'_j(\bar{x})$ , so that (8) is satisfied, and to replace the coefficients  $\alpha'_j$  by  $\alpha_j$  which will satisfy  $\alpha_j = \alpha'_j / s$  for the solution.

Therefore, consider a probability density function, of the following form

$$f(\bar{x}) = \sum_{j=1}^m \alpha_j \psi_j(\bar{x}) ,$$

where the  $\psi_j(\bar{x})$ ,  $j = 1, \dots, m$ , satisfy (8) and  $\alpha_j$  are the weighting coefficients of the mixture of these density functions. It is obvious that for a random vector  $\bar{x}$  that satisfies the distribution  $f(\bar{x})$ ,  $\alpha_j$  can be considered as the prior probability by which the vector will be attributed to component  $j$  of the mixture – satisfying

$$\sum_{j=1}^m \alpha_j = 1 . \quad (9)$$

The function  $\psi_j$  can be considered as the probability density of  $\bar{x}$ , conditioned by the knowledge that the vector is attributed to component  $j$  of the mixture, i.e.:  $\psi_j(\bar{x}) = f(\bar{x}|j)$ .

When a sample  $\{\bar{x}_1, \dots, \bar{x}_p\}$  of independent observations of the random vector  $\bar{x}$  is given, together with a predetermined set of probability density functions  $\psi_j(\bar{x})$ ,  $j = 1, \dots, m$ , of the mixture, the weighting coefficients  $\alpha_j$  can be estimated using a maximum likelihood approach. It is obvious that the likelihood to be maximised is

$$L = \prod_{l=1}^p \sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l) ,$$

or equivalently, the log-likelihood to be maximised is

$$\log L = \sum_{l=1}^p \log \sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l) .$$

The maximisation for the values of  $\hat{\alpha}_j$  must be subject to the constraint (9), and, therefore, involves the use of a Lagrange multiplier  $\lambda$ . We thus search for the solution of

$$\frac{\partial}{\partial \hat{\alpha}_i} \left( \sum_{l=1}^p \log \sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l) + \lambda \sum_{j=1}^m \hat{\alpha}_j - \lambda \right) = 0 , \quad \forall i \in \{1, \dots, m\} ,$$

which yields

$$-\lambda = \sum_{l=1}^p \frac{\psi_i(\bar{x}_l)}{\sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l)} , \quad \forall i \in \{1, \dots, m\} .$$

Multiplying both sides of this equation with  $\hat{\alpha}_i$ , we obtain a set of equations which allow us to calculate the value of  $\lambda$ :

$$-\lambda \hat{\alpha}_i = \sum_{l=1}^p \frac{\hat{\alpha}_i \psi_i(\bar{x}_l)}{\sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l)} , \quad \forall i \in \{1, \dots, m\} . \quad (10)$$

Indeed, summing both sides of these equations over index  $i \in \{1, \dots, m\}$ , interchanging the order of the summations  $\sum_i$  and  $\sum_l$  in the right hand side, and using (9), gives:  $-\lambda = \sum_{l=1}^p 1 = p$ . Substituting this result in (10) finally yields a set of equations from which the values of  $\hat{\alpha}_j$  can be solved. We have

$$\hat{\alpha}_i = \frac{1}{p} \sum_{l=1}^p \frac{\hat{\alpha}_i \psi_i(\bar{x}_l)}{\sum_{j=1}^m \hat{\alpha}_j \psi_j(\bar{x}_l)} , \quad \forall i \in \{1, \dots, m\} . \quad (11)$$

This set of equations is immediately formulated in a form, appropriate for the application of a recursive solution procedure. In such approach, one starts

with a tentative set of coefficients  $\hat{\alpha}_i$ ,  $i \in \{1, \dots, m\}$ , and plugs it in the right hand side of (11), yielding a new set of estimates for the coefficients  $\hat{\alpha}_i$ . This is repeated till a convergence criterion is satisfied. When the starting values of the coefficients are all positive, it is obvious from (11), that they remain positive during the whole procedure, since the functions  $\psi_i$  are also positive. It is also evident from (11) that the sum of the coefficients  $\hat{\alpha}_i$  is one. For this reason and because of (8), the proposed ML solution will satisfy all properties of a PDF.

An interesting set of equations, similar to the equations derived by Duda and Hart ([2], pp. 192, 193), can be derived from (11), by replacing the functions  $\psi_j(\bar{x})$  with their expression as conditional probabilities, namely  $f(\bar{x}|j)$ , and using the Bayes theorem:

$$\hat{\alpha}_i = \frac{1}{p} \sum_{l=1}^p P(i|\bar{x}_l) .$$

We indeed see that the coefficient  $\hat{\alpha}_i$  can be considered as the mean posterior probability to attribute the observed random vectors  $\bar{x}_l$  to component  $i$  of the mixture probability density model.

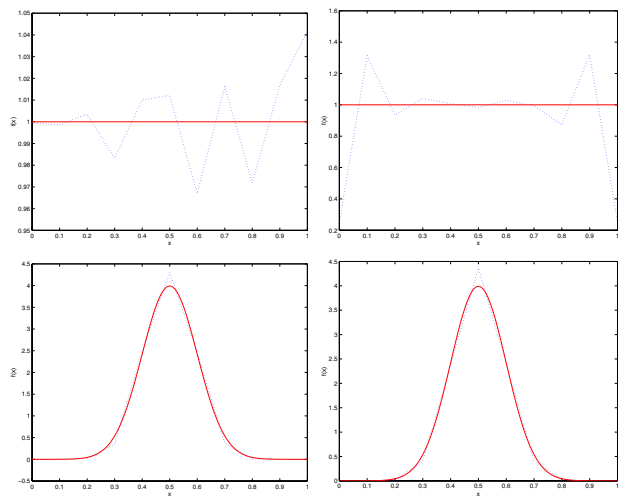
### 3 Some Numerical Experiments

The mathematical models have been implemented in Matlab. Some numerical experiments have been performed to validate the developed software technically, and to observe the behaviour of both the LS and the ML approach. One of the experiments consisted of estimating the PDFs of two univariate distributions. The first distribution is uniform in an interval  $[0, 1]$ . The second distribution behaves like a Gaussian distribution in an interval  $[0, 1]$  and is zero elsewhere. Fig. 1 shows the results of the experiments. In [1], we already reported the decrease in quality of the results when there is some mismatch between the estimated PDF and the approximation model. The figure shows that for the LS technique the values of the approximating function may be indeed negative in the neighbourhood of rapid changes. As predicted theoretically, the approximating function obtained from the ML approach continues to behave well.

### 4 Application in Digital Subtraction Radiography

Digital subtraction radiography (DSR) is a potentially sensitive method for revealing subtle changes in radiolucency between radiographs acquired separated in time. The power of the DSR method stems from its ability to remove so-called structural noise, arising from the invariant structures in the images, leading to a distinctive improvement in the visual acuity of real changes.

The particular application of DSR that we consider here as an illustration of the new method for piecewise linear approximation of a PDF is that of intra-oral radiography, and more specifically its application to the detection of approximal caries. Basically, dental caries is a slowly progressing demineralisation of the tooth surface that starts at the surface, and gradually penetrates the



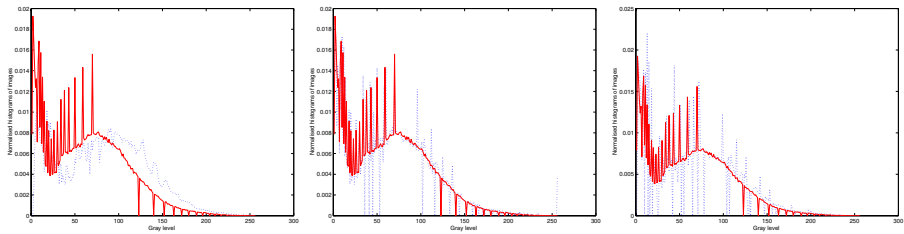
**Fig. 1.** Graphical representation of the results of the numerical experiments. Solid lines correspond to the real PDF. Top row: results for a uniform distribution, bottom row: results for a (partially) gaussian distribution, left column: results for the LS method, right column: results for the ML method

tooth. Given the treatment ramifications of advanced caries, detection in its early development stage is of prime importance. However, common radiographic examination has been found to yield an insufficient sensitivity for the early detection of approximal caries. Given this observation, DSR is investigated as a more accurate diagnostic method.

Decisive to the success of the DSR method, however, is the ability with which the exposure geometry and the development conditions can be reproduced. Whereas the first requirement can be met by the use of mechanical stabilisation devices or by employing mathematical methods of retrospective geometric correction, changes in the exposure and development conditions necessarily call for a numerical contrast correction procedure. This involves the transformation of the gray value histograms such as to resemble each other as closely as possible.

The standard method of contrast correction used in intra-oral DSR is that, proposed by Ruttiman and co-workers [8], which finds the optimal transformation by equating the cumulative distributions of the respective histograms. Actually, this method has been proposed as a more consistent approach to the problem of contrast correction, compared to an earlier described parametric method [9], based on matching the first and second order moments of the respective distributions.

More recently, a particular interesting method has been suggested by Bidasaria [10], in which the original gray values in the images are randomised in



**Fig. 2.** Plots of desired histogram (solid line) with, (a) Actual starting histogram, (b) Matched histogram from piecewise linear approximation, (c) Matched histogram using the method of Ruttiman

the discretisation intervals, to obtain a piecewise constant approximation of the histograms. Histogram matching then follows immediately.

In our approach, the histogram of image 1, say, is transformed into that of image 2, by first decomposing the histograms into the basis set of triangular functions characterising the PL approximation, prior to using the method of direct histogram specification (DHS) [11] via the uniform distribution. Upon a suitable choice of the points at which the cumulative distribution function (CDF) of image 2 is evaluated, the use of a continuous representation allows to circumvent the explicit inversion of the transformation to the uniform distribution of histogram 2. Our approach, consisting of a PL approximation, differs fundamentally from that of Bidasaria [10] in which a step approximation of the histogram is proposed. In the preliminary results of this paper, we have shown as in Fig. 2 that our approach yields results comparable to those obtained from the method of Ruttiman et al. [8]. Our approach is a first step towards the use of alternative representations of histograms, as found in [12,13].

## 5 Discussion and Conclusions

In [1], we introduced the basic concepts and notations for a Piecewise Multilinear (PML) approximation of probability density functions. We showed how to formulate this model, which is basically an interpolation model, as a weighted sum of basis functions, where the weights are the model parameters. We also proposed a solution methodology for estimating the model parameters from a representative learning set of patterns, based on a least squares (LS) fitting criterion.

For a broad class of models that includes the PML model, we show in this paper that the optimal approximation of the PDF by the weighted sum of basis functions, as well as the estimate of this approximation from a learning set of patterns, satisfy the property that their integral over the definition domain equals unity, thus meeting a basic property of a PDF (Theorems 1 and 2). To cope with the problem that another property of PDFs — namely positivity — is



not always satisfied by the LS fitting solution, we introduce another approach, which is based on a Maximum Likelihood (ML) criterion. The ML estimate satisfies all properties of a PDF and hence, can be used in applications where these properties are required. Subsequently, we have demonstrated an application of the PL approximation to the problem of contrast correction in DSR.

## Acknowledgement

We thank Mandy Runge for her assistance with the preparation of Fig. 2.

## References

1. Edgard Nyssen, Luc Van Kempen, and Hichem Sahli. Pattern classification based on a piecewise multi-linear model for the class probability densities. In *Advances in Pattern Recognition — proceedings SSPR2000 and SPR2000*, pages 501–510, 2000. 752, 753, 757, 759
2. Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973. 752, 753, 757
3. Julius T. Tou and Raphael C. Gonzales. *Pattern Recognition Principles*. Addison Wesley Publishing Company, 1979. 752
4. Robert Schalkoff. *Pattern Recognition — Statistical, Structural and Neural Approaches*. John Wiley & Sons, 1992. 752
5. Fang Sun, Shin'ichiro Omachi, and Hiroto Aso. An algorithm for estimating mixture distributions of high dimensional vectors and its application to character recognition. In *Proc. 11th Scandinavian Conference on Image Analysis*, pages 267–274, 1999. 752
6. David L. Donoho, Iain M. Johnstone, Gérard Kerkycharian, and Dominique Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2):508–539, 1996. 752
7. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm. *J. Royal Statistical Society*, 39:1–38, 1977. 753
8. U. E. Ruttimann, R. L. Webber, and E. Schmidt. A robust digital method for film contrast correction in subtraction radiography. *J. Periodont. Res.*, 21:486–495, 1986. 758, 759
9. U. E. Ruttimann, T. Okano, H.-G. Gröndahl, K. Gröndahl, and R. L. Webber. Exposure geometry and film contrast differences as bases for incomplete cancellation of irrelevant structures in dental subtraction radiography. *Proc. SPIE*, 314:372–377, 1981. 758
10. H. B. Bidasaria. A method for almost exact histogram matching for 2 digitized images. *Computer Graphics and Image Processing*, 34, 1986. 758, 759
11. Rafael C. Gonzalez and Paul Wintz. *Digital image processing*. Addison-Wesley Publishing Company, Amsterdam, 1987. 759
12. R. Morandi and P. Constantini. Piecewise monotone quadratic histosplines. *SIAM J. Stat. Comput.*, 10:397–406, 1989. 759
13. J. W. Schmidt, W. Heß, and T. Nordheim. Shape preserving histopolation using rational quadratic splines. *Comput.*, 44:245–258, 1990. 759