# Chapter 2

# Discrete LPA

The idea of local smoothing and local approximation is so natural that it is not surprising it has appeared in many branches of science. Citing [145] we mention early works in statistics using local polynomials by the Italian meteorologist Schiaparelli (1866) and the Danish actuary Gram (1879) (famous for developing the Gram-Schmidt procedure for the orthogonalization of vectors).

In the 1960s and 1970s the idea became the subject of intensive theoretical study and applications, in statistics by Nadaraya [158], Watson [230], Parzen [175], and Stone [215], and in engineering sciences by Brown [19], Savitzky and Golay [196], Petersen [179], Katkovnik [98], [99], [101], and Cleveland [29].

The local polynomial approximation as a tool appears in different modifications and under different names, such as moving (sliding, windowed) least square, Savitzky-Golay filter, reproducing kernels, and moment filters. We prefer the term LPA with a reference to publications on nonparametric estimation in mathematical statistics where the advanced development of this technique can be seen.

In this chapter the discrete local approximation is presented in a general multivariate form. In the introductory section (Sec. 2.1), we discuss an observation model and multi-index notation for multivariate data, signals, and estimators.

Section 2.2 starts with the basic ideas of the LPA presented initially for 2D signals typical for image processing. The window function, the order of the LPA model, and the scaling of the estimates are considered in detail. Further, the LPA is presented for the general multivariate case with estimates for smoothing and differentiation. Estimators for the derivatives of arbitrary orders are derived. Examples of 1D smoothers demonstrate that the scale (window size) parameter selection is of importance.

The LPA estimates are given in kernel form in Sec. 2.3. The polynomial smoothness of the kernels and their properties are characterized by the vanishing moment conditions.

The LPA can be treated as a design method for linear smoothing and differentiating filters. Links of the LPA with the nonparametric regression concepts are clarified in Sec. 2.4, and the LPA for interpolation is discussed briefly in Sec. 2.5.

## 2.1 Introduction

### 2.1.1 Observation modeling

Since the LPA is a method that is valid for signals of any dimensionality, we mainly use a multidimensional notation because it is universal and convenient. For instance, we say that the signal $y(x)$ depends on $x \in \mathbb{R}^d$, i.e., the argument $x$ is a $d$-dimensional vector, and we use it with $d = 1$ for scalar signals in time and with $d = 2$ for 2D images. The multidimensional notation allows us to present results in a general form valid for both scalar time and 2D image and signal processing.

The symbol $d$ is reserved for the dimensionality of the argument-variable $x$. Thus, $x$ is a vector with elements $x_1, \ldots, x_d$.

Suppose that we are given observations of $y$ in the form $y_s = y(X_s)$, $s = 1, \ldots, n$. Here $X_s$ stays for a location of the $s$th observation, and the coordinates of this location are

$$X_s = [x_1(s), \ldots, x_d(s)].$$

The integers $s$ number observations and their locations.

Noisy observations with additive noise can be given in the following standard model commonly used for image and signal processing:

$$z_s = y(X_s) + \varepsilon_s, s = 1, \ldots, n, \tag{2.1}$$

where the additive noise $\varepsilon_s$ is an error of the $s$th experiment usually assumed to be random, zero-mean independent for different $s$ with $E\{\varepsilon_s\} = 0$, $E\{\varepsilon_s^2\} = \sigma^2$.

Equation (2.1) is a *direct observation model* where a signal $y$ is measured directly and the additive noise only defines a data degradation.

Suppose we are able to observe $(y \circledast v)(x)$, where $v$ is a *blurring kernel* or *point spread function* (PSF) of a linear convolution. The blurring phenomenon modeled by the PSF $v$ (continuous or discrete) is very evident in many signal and image applications. Moreover, we assume that the data are noisy, so that we observe $z_s$ given by

$$z_s = (y \circledast v)(X_s) + \varepsilon_s. \tag{2.2}$$

This equation is an *indirect observation model* with a signal $y$ measured not directly but after some transformation.

A data degradation is defined by this transformation as well as by the additive noise. It is assumed in image processing that all signals in Eqs. (2.1) and (2.2) are defined on a 2D rectangular *regular grid*:

$$X = \{X_s : s = 1, \ldots, n\}$$

with pixel (grid node) coordinates

$$X_s = [x_1(s), x_2(s)]. \tag{2.3}$$

In this notation the image pixels are numbered by $s$ taking values from 1 through $n$. The coordinates of these pixels can be given explicitly as

$$X = \{k_1\Delta_1, k_2\Delta_2 : k_1 = 1, \ldots, n_1, k_2 = 1, \ldots, n_2\}, \qquad n = n_1 n_2, \tag{2.4}$$

where $\Delta_1$ and $\Delta_2$ are the sampling intervals for $x_1$ and $x_2$, respectively.

In signal processing literature, square brackets are sometimes used for discrete arguments to differentiate them from continuous arguments. For example, the notation $y(k_1\Delta_1, k_2\Delta_2)$ is for a continuous argument function and $y[k_1, k_2]$ is for a discrete one with

$$y(k_1\Delta_1, k_2\Delta_2) = y[k_1, k_2]. \tag{2.5}$$

For this discrete argument $y$ we also can use the general notation $y(X_s)$ provided that $X_s = (k_1\Delta_1, k_2\Delta_2)$.

The LPA considered in the book is applicable to observations given on irregular grids, when the coordinates $X_s$ and the distances between $X_s$ can be arbitrary. In image processing this sort of situation appears when data are destroyed or lost. Figure 2.1 illustrates the cases of regular and irregular grids for 2D data.

### 2.1.2 Classes of signals

A basic objective of data processing is to reconstruct (estimate) $y$ from the noisy observations $\{z_s\}$ with pointwise errors that are as small as possible. It is assumed that $y$ is an unknown deterministic. In comparison with stochastic models of $y$, this means that the main intention is to obtain the best results for every realization of $y$, even if they are generated by a probabilistic phenomenon.

The first category is that of *parametric models* $y(x, \theta)$, where $\theta$ is a vector of unknown parameters. These models seek a signal $y$ with parameter $\theta$ that gives an approximation of the relation between $x$ and $y$. We predict $y$ based on the argument variable $x$ assuming that $\theta$ is given. The estimation is concentrated around methods for finding $\theta$ from observations

$$z_s = y(X_s, \theta) + \varepsilon_s, \qquad s = 1, \ldots, n.$$

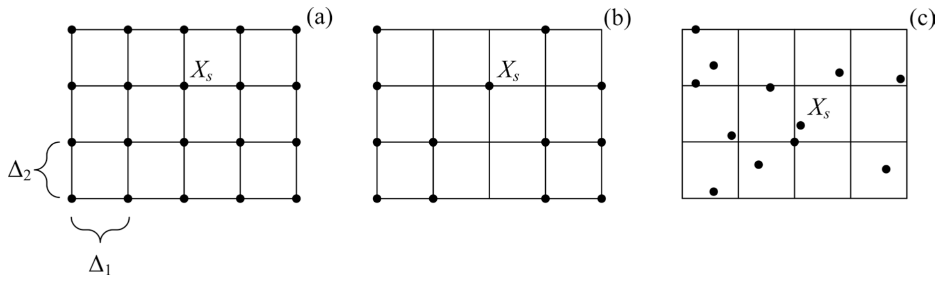


**Figure 2.1** Types of data grids: (a) regular with sampling intervals $\Delta_1$ and $\Delta_2$ on the arguments $x_1$ and $x_2$; (b) regular with missing data; and (c) irregular with varying intervals between observations $X_s$.

The second category is that of *nonparametric models*, which assume that a parametric representation of $y$ with a reasonably small number of parameters $\theta$ is unknown or does not exist. This $y$ can be quite arbitrary, say, with discontinuities and varying curvature. Examples of nonparametric and parametric curves (linear $y(x,\theta) = \theta_1 + \theta_2 x$ and harmonic $y(x,\theta) = \theta_1 \cos(\theta_2 x + \theta_3) + \theta_4$) are shown in Fig. 2.2.

In this book we deal with the nonparametric category of signals $y$. Usually it is assumed that there is some continuous argument signal $y(x)$, and the observed $y(X_s)$ are sampled values of this continuous signal. Different classes of continuous argument signals can be exploited. We restrict our analysis to the class $\mathbb{C}^\alpha$ of continuous differentiable signals $y$ of $d$ variables with bounded $\alpha$th-order derivatives, which are defined mathematically as follows

$$\mathbb{C}^\alpha = \left\{ y : \max_{r_1+\cdots+r_d=\alpha} \left| \partial_{x_1}^{r_1} \ldots . \partial_{x_d}^{r_d} y(x) \right| = L_\alpha(x) \leq \overline{L}_\alpha, \forall x \in \mathbb{R}^d \right\}, \tag{2.6}$$

where $L_\alpha(x)$ and $\overline{L}_\alpha$ are finite and $\alpha = r_1 + \cdots + r_d$ is the order of the derivative used in this definition.

The class $\mathbb{C}^\alpha$ is composed of signals having bounded all possible derivatives of the order $\alpha$.

Let $d = 1$. Then the class $\mathbb{C}^\alpha$ is defined as

$$\mathbb{C}^\alpha = \{ y : |\partial_{x_1}^\alpha y(x_1)| = L_\alpha(x_1) \leq \overline{L}_\alpha, \forall x_1 \in \mathbb{R} \}.$$

The class includes all signals with the bounded $\alpha$th-order derivative.

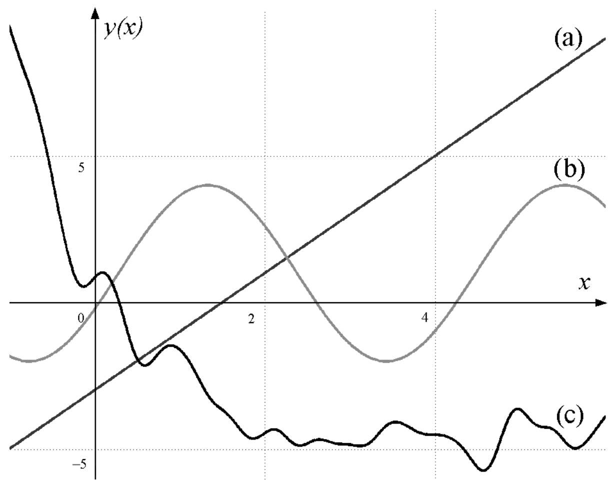


**Figure 2.2** Parametric and nonparametric models: (a) linear parametric model, (b) cosine parametric model, and (c) nonparametric function with unknown parametric representation.

Let $d = 2$. Then

$$\mathbb{C}^{\alpha} = \left\{ y : \max_{r_1+r_2=\alpha} \left| \partial_{x_1}^{r_1} \partial_{x_2}^{r_2} y(x_1, x_2) \right| = L_{\alpha}(x_1, x_2) \leq \overline{L}_{\alpha}, \forall (x_1, x_2) \in \mathbb{R}^2 \right\}.$$

In this definition, the maximum for each $x = (x_1, x_2)$ is calculated over all derivatives of the order $\alpha = r_1 + r_2$. For $\alpha = 1$, it assumes that

$$\max(|\partial_{x_1} y(x)|, |\partial_{x_2} y(x)|) = L_1(x),$$

while for $\alpha = 2$ it means

$$\max(|\partial_{x_1}^2 y(x)|, |\partial_{x_2}^2 y(x)|, |\partial_{x_1} \partial_{x_2} y(x)|) = L_2(x).$$

Definition (2.6) is global because the conditions hold for all $x$. For images where singularities and discontinuities of $y$ carry the most important image features, a local smoothness of $y$ is more realistic. Therefore, the class $\mathbb{C}^{\alpha}$ can be replaced by

$$\mathbb{C}_A^{\alpha} = \left\{ y : \max_{r_1+\cdots+r_d=\alpha} \left| \partial_{x_1}^{r_1} \ldots \partial_{x_d}^{r_d} y(x) \right| = L_{\alpha}(x) \leq \overline{L}_{\alpha}, \forall x \in A \subset \mathbb{R}^d \right\}, \quad (2.7)$$

where the argument $x$ is restricted to a subset $A$.

The following piecewise smooth models are useful for many applications.

Let an area where $y$ is defined be separated into $q$ regions $A_i$, $i = 1, \ldots, q$, and each of these regions be a connected set with edges (boundaries $G_i$). The signal $y$ is assumed to be smooth differentiable within $A_i$:

$$y(x) = \sum_{i=1}^{q} y_i(x) \mathbf{1}[x \in A_i], \quad (2.8)$$

where $\mathbf{1}[x \in A_i]$ is an indicator of the region $A_i$, $\mathbf{1}[x \in A_i] = 1$ if $x \in A_i$ and zero otherwise. More specifically, $y_i$ belongs to a class of continuous differentiable signals $\mathbb{C}_{A_i}^{\alpha}$. The parameters $\alpha$, $L_{\alpha}(x)$, and $\overline{L}_{\alpha}$ can be different for the different $y_i$ in Eq. (2.8).

The piecewise constant $y$,

$$y(x) = \sum_{i=1}^{q} a_i \mathbf{1}[x \in A_i], \quad (2.9)$$

is a particular case of Eq. (2.8) with constant values within each region $A_i$. Figure 2.3 shows an example of a piecewise constant $y$ defined according to Eq. (2.9) as well as the corresponding regions $A_i$.

In the models (2.8) and (2.9) $y_i(x)$, and $a_i$, as well as the regions $A_i$ are usually unknown. The boundaries $G_i$ define change points of the piecewise smooth $y$ in Eq. (2.8). The estimation of $y$ can be produced in different ways. One possible approach deals with a two-stage procedure that includes estimating the boundaries $G_i$ at the first stage, which defines the regions $A_i$. The second stage is a parametric or nonparametric fitting $y_i$ in $A_i$.

Another approach is connected with the concept of spatially adaptive estimation. In this context, change points can be viewed as a sort of anisotropic behavior

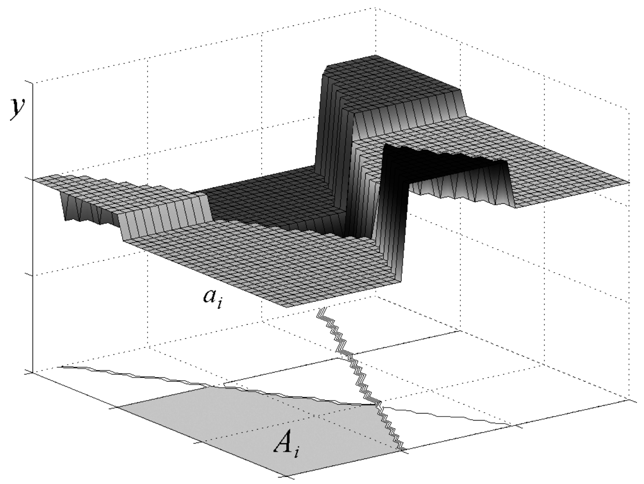


**Figure 2.3** A piecewise constant 2D function.

of the estimated signal. One may therefore apply the same procedure for all $x$, for instance, nonlinear wavelet, ridgelet, or curvelet estimators, and the analysis focuses on the quality estimation when change-points are incorporated into the model. With this approach, the main intention is to estimate the signal but not the locations of the change-points, which are treated as elements of the signal surface.

In this book we will follow the second approach. The objective is to develop a method that simultaneously adapts to the anisotropic smoothness of the estimated curve and is also sensitive to the discontinuities in the curves and their derivatives.

In general, conditions (2.6) and (2.7) define nonparametric signals that do not have parametric representations. The following important class of parametric $y$ can be interpreted as a special case of the nonparametric class $\mathbb{C}^{\alpha}$.

Let $L_{m+1}(x) \equiv 0$ for $x \in \mathbb{R}^d$. Then $y$ is polynomial of the degree $m$. Indeed, the condition $L_{m+1}(x) \equiv 0$ means that $y$ is a solution of the set of differential equations

$$\partial_{x_1}^{r_1} \dots . \partial_{x_d}^{r_d} y(x) = 0$$

for all $r_i$ such that $r_1 + \cdots + r_d = m + 1$, and this solution is a polynomial. For this class of polynomials we use the notation

$$\mathcal{P}^m = \{y : \partial_{x_1}^{r_1} \dots . \partial_{x_d}^{r_d} y(x) = 0, r_1 + \cdots + r_d = m + 1, \forall x \in \mathbb{R}^d\}, \qquad (2.10)$$

where $m$ stays for the power of the polynomial. It is obvious that $\mathcal{P}^m \subset \mathbb{C}^{m+1}$ with $\overline{L}_{m+1} = 0$.

### 2.1.3 Multi-index notation

*Multi-index notation* is used to shorten expressions that contain many indices. Let $x$ be a $d$-dimensional vector, $x \in \mathbb{R}^d$, and $y$ be a scalar function of $x$,