Signal Processing 108 (2015) 576–588

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Slope estimation in noisy piecewise linear functions ☆

Atul Ingle [a,b,*,1], James Bucklew [a], William Sethares [a], Tomy Varghese [b,a]

[a] Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA
[b] Department of Medical Physics, University of Wisconsin-Madison, Madison, WI 53705, USA

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form
26 June 2014
Accepted 2 October 2014
Available online 30 October 2014

Keywords:
Piecewise linear function
MAP estimation
Dynamic programming optimization
EM algorithm
Alternating maximization

ABSTRACT

This paper discusses the development of a slope estimation algorithm called MAPSlope for piecewise linear data that is corrupted by Gaussian noise. The number and locations of slope change points (also known as breakpoints) are assumed to be unknown a priori though it is assumed that the possible range of slope values lies within known bounds. A stochastic hidden Markov model that is general enough to encompass real world sources of piecewise linear data is used to model the transitions between slope values and the problem of slope estimation is addressed using a Bayesian maximum a posteriori approach. The set of possible slope values is discretized, enabling the design of a dynamic programming algorithm for posterior density maximization. Numerical simulations are used to justify choice of a reasonable number of quantization levels and also to analyze mean squared error performance of the proposed algorithm. An alternating maximization algorithm is proposed for estimation of unknown model parameters and a convergence result for the method is provided. Finally, results using data from political science, finance and medical imaging applications are presented to demonstrate the practical utility of this procedure.

1. Introduction

The need for piecewise linear regression arises in many different fields, as diverse as biology, geology, and the social sciences. This paper addresses the problem of direct estimation of slopes from piecewise linear data. An important application of interest for this paper is ultrasound shear wave elastography, where ultrasonic echoes are used to track the motion of an externally generated mechanical shear wave pulse traveling through multiple tissue interfaces [19]. The time of arrival of this shear wave pulse is recorded as a function of spatial coordinates in the ultrasound imaging plane and the reciprocal of the slope of this function gives an estimate of the speed of the wave. Breakpoints (where the slope changes) indicate tissue interfaces. These estimates are useful from a clinical perspective because they provide a way to quantify mechanical properties of tissue, thereby adding value to subjective judgments about the location and size of cancerous tumors.

A similar issue in larger spatial dimensions occurs in seismology where the time of arrival of seismic waves is tracked at different locations relative to the epicenter of an earthquake. The velocity of these waves provides information about the mechanical properties of the geological medium. Piecewise linear data also occurs in the study of flow of soil through water streams and is referred to as bedload data [20].

☆ This work was supported in part by NIH-NCI Grants R01CA112192-S103 and R01CA112192-05.

* Corresponding author at: Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53705, USA.
E-mail addresses: ingle@wisc.edu (A. Ingle), bucklew@engr.wisc.edu (J. Bucklew), sethares@ece.wisc.edu (W. Sethares), tvarghese@wisc.edu (T. Varghese).

[1] Sample MATLAB code is available on the author's website at http://homepages.cae.wisc.edu/~ingle/mapslope.html.

http://dx.doi.org/10.1016/j.sigpro.2014.10.003
0165-1684/© 2014 Elsevier B.V. All rights reserved.

1.1. Data model

Assume that the piecewise linear data is generated by the following discrete time hidden Markov model (HMM). The underlying (unknown) function takes on values Z_n at each discrete index $1 \le n \le N$. This function value is obtained by accumulating slope values S_k up to the time index n. Zero mean Gaussian noise with variance σ^2 is added to each running sum resulting in the observed function value X_n. Also, suppose that for any n, the probability of maintaining the previous slope value is p and the probability of transitioning into a new slope value is $1-p$. These relations can be written mathematically as follows:

$$\begin{aligned} Z_0 &= 0 \quad \text{with probability } 1, \\ Z_n &= Z_{n-1} + S_n, \\ X_n &= Z_n + w_n \end{aligned} \tag{1}$$

for $n = 1, \ldots, N$ where $w_n \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$. A Markov structure is imposed on the slope values as follows:

$$S_n = \begin{cases} S_{n-1} & \text{with probability } p \\ U_n & \text{with probability } 1-p \end{cases}$$

for $n = 2, \ldots, N$ where $U_n \sim \mathcal{U}(\{0, 1/(M-1), \ldots, (M-2)/(M-1), 1\} \setminus \{S_{n-1}\})$ denotes a discrete uniform random variable taking on one of $M-1$ possible slope values and the initial slope value is drawn uniformly as $S_1 \sim \mathcal{U}(\{0, 1/(M-1), \ldots, (M-2)/(M-1), 1\})$.

Another implicit assumption is that the slopes can take on values on a closed bounded interval $[s_l, s_u]$ with upper and lower limits $0 < s_l < s_u < \infty$ known a priori. For instance, in the ultrasound-based wave tracking application, the values of s_l and s_u can be obtained from the underlying physics which dictates that such mechanical waves travel with speeds between 0.5 and 10 m/s in homogeneous tissue. With the knowledge of s_l and s_u, the given data vector can be translated and rescaled so that all slope values lie in the interval [0,1]. Hence, without loss of generality, it suffices to design a slope estimation algorithm that operates with a finite set of slopes $\mathcal{S} = \{0, 1/(M-1), \ldots, (M-2)/(M-1), 1\}$. Intuitively, this quantization step is justified because in the presence of noise it is impossible to detect the difference between slope values that differ only slightly.

1.2. Main contributions

The main contributions of this paper are as follows:

(a) a hidden Markov model formulation of the slope estimation problem that is general enough to encompass different applications;
(b) a procedure for maximum a posteriori (MAP) estimation of slopes from this Markov model;
(c) a dynamic programming routine on a linearly growing trellis for fast MAP estimation;
(d) a mean squared error (MSE) optimality analysis of this routine via simulations and a comparison with reasonable upper and lower bounds;
(e) an alternating maximization algorithm that alternately maximizes an objective function with respect to the unknown sequence of slope values and unknown model parameters to jointly estimate both of them from data;
(f) a comparison of the performance of this algorithm with other methods in literature applied to real world data.

1.3. Related work

In many real world applications, the local slope values of an observed noisy function have interesting physical interpretations. Most of the existing methods do not directly address slope estimation; rather, they attempt to fit a model to the data. For instance, standard regression or spline-based methods can be used to fit a smooth function to the data and local slopes can be estimated from this fit. However, even if the function-fitting algorithm generates optimal fits (according to a cost function such as the minimum MSE), there is no guarantee that the local slope estimates obtained from this fit are themselves optimal. This paper bypasses the need for such post-processing by directly estimating the slopes and breakpoints. This is particularly useful when the slopes correspond directly to the variables of interest and the breakpoints correspond to where those variables change.

The topic of slope estimation from noisy data is quite old; an early paper can be traced back to 1964 where the popular Savitzky–Golay differentiator [10] was introduced. Their main idea is to use a locally windowed least squares fit to estimate the slope at each data sample, where the window coefficients are chosen to satisfy a certain frequency response that mimics a high pass filter together with some level of noise averaging. Another similar technique that is used in statistics is called locally weighted least squares regression (LOWESS) [22]. However, these methods undesirably smooth out the breakpoint locations in when data has sharp transitions or jumps. In contrast, the algorithm in the present paper prevents blurring the transitions by explicitly allowing for sharp slope transitions using a Markov model.

In some situations, the raw data can be massaged using a preprocessing step so that it becomes piecewise linear. The simplest example is the case of piecewise constant data — the running sum (integral) of such data yields a piecewise linear function. Ratkovic and Eng [21] discuss a statistical spline fitting approach combined with the Bayesian information criterion (BIC) to detect abrupt transitions in political approval ratings. Data from their paper is used in Section 6.1. As a special case, their method can be applied when function values stay almost constant over long intervals and occasionally shift to a new value. In another application, Bai and Perron [16] use statistical regression techniques to detect multiple regime shifts in interest rate data. The algorithm developed in the present paper provides comparable numerical performance as the Bai–Perron algorithm as shown in Section 6.2.

Closely related problems of piecewise linear regression for noisy data have been addressed over the years. For example, an early paper by Hudson [1] focuses on a technique to obtain