ESTIMATING PIECEWISE LINEAR MODELS USING COMBINATORIAL OPTIMIZATION TECHNIQUES

Marco Mattavelli§ and Edoardo Amaldi† *

§*Signal Processing Laboratory, Swiss Federal Institute of Technology*
CH-1015 Lausanne, Switzerland, E-mail: marco.mattavelli@lts.de.epfl.ch
†*School of Operations Research and Center for Applied Mathematics*
Cornell University, Ithaca, NY 14853, USA, E-mail amaldi@cs.cornell.edu

ABSTRACT

A wide range of image and signal processing problems have been formulated as ill-posed linear inverse problems. Due to the importance of discontinuities and non-stationarity, piecewise linear models are a natural step towards more realistic results. Although there have been some attempts to extend classical approaches to deal with discontinuities, finding at the same time the piecewise decomposition and the corresponding model parameters remains a major challenge. A new approach based on partitioning inconsistent linear systems into a minimum number of consistent subsystems (MIN PCS) is proposed for solving ill-posed problems whose formulation as linear inverse problems with discrete data fails to take into account discontinuities. In spite of the NP-hardness of MIN PCS, satisfactory approximate solutions can be obtained using simple but effective variants of an algorithm which has been extensively studied in the artificial neural network literature. Our approach presents various advantages compared to classical alternatives, including a wider range of applicability and a lower computational complexity.

1. INTRODUCTION

Many image and signal processing problems center around extracting high-level information from low-level or raw data. Despite the crucial role that discontinuities and non-stationarity play in such problems, standard modeling techniques are based on continuous functions and operators. This is the case of regularization theory (see [1]), which has been proposed as a unified and standard way of finding approximate solutions to ill-posed problems. The idea is to introduce *a priori* information to deal with noise and to cope with unreliable or missing data. This is achieved either by using variational principles that impose constraints on the admissible solutions or by making statistical assumptions on the solution space (see [2, 3] and the included references). The *smoothness* constraint is a typical example of widely used *a priori* information. While such approaches based on continuous operators face serious limitations, considering discontinuous functions and operators explicitly leads to a different class of models and methods which pertain to Discrete Mathematics. Similar discussions are also true for the various techniques of adaptive signal processing [4, 5].

Since practical linear inverse problems are always made discrete, they reduce to linear systems. When linear models are too simple to deal with the full complexity of the problem at hand (the corresponding linear system is inconsistent), a natural step is to consider piecewise linear models that are able to model more complex phenomena but are still simple enough to estimate locally because of linearity. But determining the structure of the data (detecting discontinuities) as well as estimating the parameters of the model turns out to be a major challenge. Indeed, structure and parameter estimation is known to be a *chicken and egg problem.*

A simple approach consists in breaking down the estimation of piecewise linear models into two distinct stages. First one tries to determine the underlying domain decomposition (partition) using some clustering methods and then one estimates the parameters associated with each component using regression or robust regression techniques. However, generally, the number of components has to be guessed in advance, the computational requirements are very high and the clustering procedures do not take into account the type of model (linear) used for each component. Although the Hough Transform [6, 7] (HT) can in principle solve *ill-posed* inverse problems without *a priori* assumptions on the data structure, the complexity and storage requirements needed to guarantee a good accuracy are prohibitive in most applications.

Thus, rather than trying to transform the problem in order to remove its *ill-posedness* (i.e. in this case the non-existence of the solution) as it is done by regularization techniques, a more natural alternative is to look for *solutions* that are partially *consistent* with the data or, more precisely, with *significant* subsets of the data. This amounts to directly taking into account discontinuities in the problem formulation. Instead of considering

*E.A. is partially funded by a Postdoctoral fellowship of the Swiss National Science Foundation.

linear models which are clearly too simple, we focus on piecewise linear ones which are able to model highly complex phenomena while remaining simple enough.

In this paper, we propose a new combinatorial optimization approach which enables estimation of piecewise linear models and, therefore, provides solutions to *ill-posed* linear problems involving discontinuities.

2. MINIMUM PARTITION INTO CONSISTENT SUBSYSTEMS (MIN PCS)

The idea is to formulate the estimation problem as that of finding a Partition of the linear system associated with the linear model into a MINimum number of Consistent Subsystems.

MIN PCS: Given a possibly inconsistent linear system

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

with a $p \times n$-matrix and a p-dimensional vector $\mathbf{b}$, find a partition of the rows of $A\mathbf{x} = \mathbf{b}$ into a minimum number of consistent subsystems.

The MIN PCS formulation is very attractive because it provides a natural way of addressing simultaneously the two fundamental issues in piecewise linear model design: domain decomposition and parameter estimation. Given any solution of MIN PCS, the partition indicates the piecewise decomposition and a solution associated with each consistent subsystem provides the parameters of the corresponding component.

According to the well-known Occam principle, we look for the "simplest" piecewise linear model consistent with the data, which is most likely to be the correct one. Here "simplicity" is measured in terms of the number of linear components.

In practice, we want to be able, like in regression techniques, to cope with modeling errors and noisy data. This is easily achieved by replacing each equation $\mathbf{a}^k\mathbf{x} = b^k$, where $\mathbf{a}^k$ is the kth row of A and b^k is the kth component of $\mathbf{b}$, with the two complementary inequalities

$$\mathbf{a}^k\mathbf{x} \leq b^k + \varepsilon \qquad \mathbf{a}^k\mathbf{x} \geq b^k - \varepsilon \tag{2}$$

where ε is the maximum admissible modeling error. If equations of the original system are expected to be affected by different noise levels, different error level settings can of course be used.

MIN PCS admits a very simple geometrical interpretation. Considering each equation as a point P_k whose coordinates are the n components of the kth row of A denoted by $\mathbf{a}^k$, $k \in \{1, 2, ..., p\}$, MIN PCS amounts to finding a minimum number of hyperplanes H_j containing all the points P_k. It easily verified that, while MIN PCS is trivial for points in general position (any partition into consistent subsystems yields a minimum size partition), it is not the case for non-pathological point distributions [8].

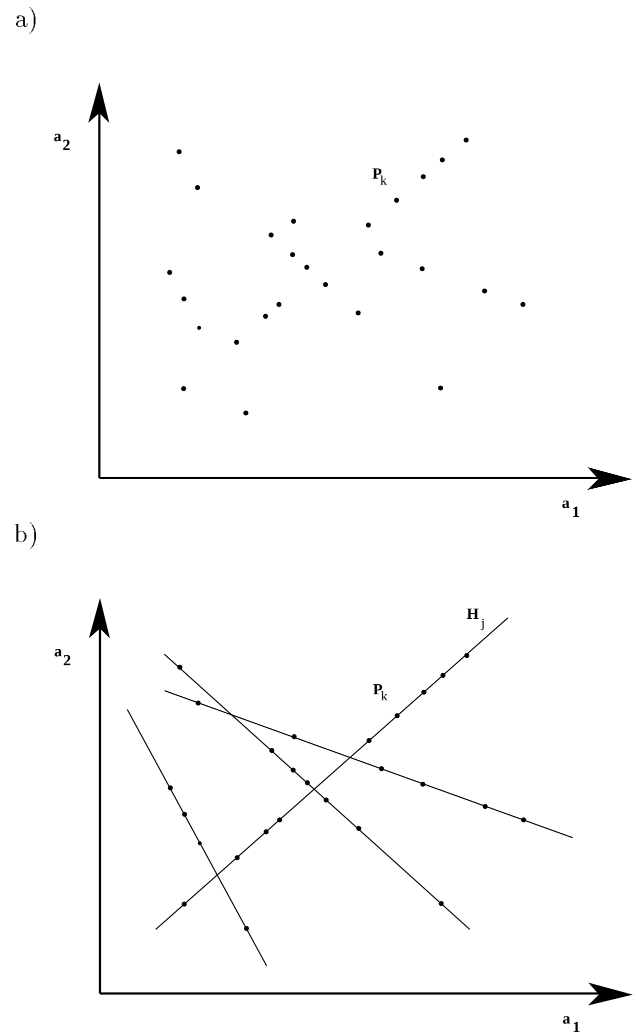

Figure 1: Geometrical interpretation of MIN PCS problem with equations in two variables. a) each point P_k corresponds to one equation, b) a graphical representation of a minimum number of hyperplanes that contains all points.

In the case of complementary inequalities such as in Eq. (2) with ϵ independent from k, MIN PCS is equivalent to finding a minimum number of *hyperslabs* of *thickness* 2ϵ containing all the points P_k with $k \in \{1, 2, ..., n\}$.

It is noteworthy that MIN PCS turns out to be a new combinatorial optimization problem (see [9] for a list of many other interesting optimization problems). To the best of our knowledge, no equivalent formulation has been proposed neither in linear system theory nor in regularization theory.

Given the relevance of MIN PCS to piecewise linear modeling, the question of its inherent complexity

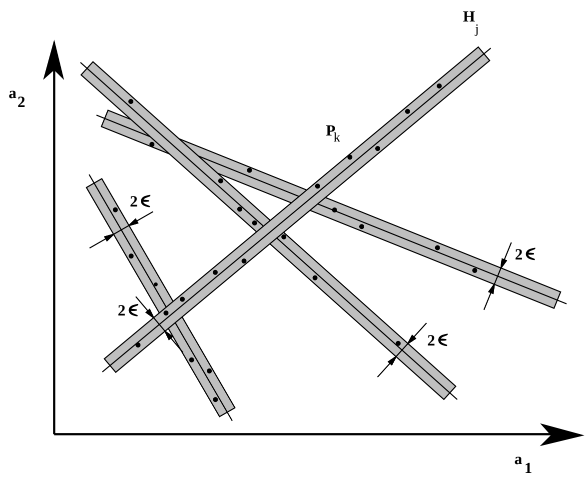

Figure 2: Geometrical interpretation of MIN PCS problem with inequalities. The problem is to find a minimum number of *hyperslabs* H_j of thickness 2ϵ containing all the points P_k.

problem arises. We show elsewhere that MIN PCS is NP-hard [8, 10], i.e. if the fundamental conjecture of complexity theory P $\neq$ NP is valid, no algorithm is guaranteed to provide an optimal solution in polynomial time. Thus, not too surprisingly, estimating a piecewise linear model turns out to be harder than finding a least mean squares solution to $A\mathbf{x} = \mathbf{b}$.

3. AN EFFICIENT HEURISTIC FOR PIECEWISE LINEAR MODEL ESTIMATION (PLIME ALGORITHM)

Although a worst-case complexity analysis provides an important insight concerning the degree of complexity of MIN PCS, it does not preclude the existence of efficient heuristics apt to finding close-to-optimal solutions within a reasonable lapse of time. Therefore, we propose a *greedy* strategy that breaks the difficult MIN PCS problem into smaller subproblems consisting of finding, iteratively, consistent subsystems with a maximum number of equations (cf. MAX FLS problem studied in [11, 12]). Clearly, eliminating repeatedly a close-to-maximum consistent subsystem until the remaining subsystem is consistent yields a partition into consistent subsystems. Since the resulting algorithm is a general method for Piecewise LInear Model Estimation, we refer to it as PLIME. This algorithm enables simultaneous estimation of the piecewise decomposition as well as the parameters of each single linear component without requiring any *a priori* assumption neither upon the number nor upon the topological *locations* of the discontinuities. Different values of the parameters ϵ_k lead to estimations having different levels of *accuracy*.

To extract maximum consistent subsystems, we use an extension of thermal variants of the perceptron procedure which have been developed and studied in the field of artificial neural networks [11, 13, 14, 15]. The algorithm can be described as follows:

Algorithm

- Problem: Given any system $A\mathbf{x} = \mathbf{b}$ and any maximum admissible error $\epsilon \geq 0$, look for an $\mathbf{x}_{max} \in \mathbb{R}^n$ such that the couple of complementary inequalities $\mathbf{a}^k\mathbf{x}_{max} \leq b^k + \epsilon$ and $\mathbf{a}^k\mathbf{x}_{max} \geq b^k - \epsilon$ is satisfied for the maximum number of indices $k \in \{1, \cdots, p\}$.
- Initialization: Take an arbitrary $\mathbf{x}_0 \in \mathbb{R}^n$, and set $c := 0$, initial temperature $t := t_0$, select a predefined number of cycles C as well as a function $\gamma(c, C)$ decreasing for increasing c and such that $\gamma(C, C) = 0$.

begin
  $i \leftarrow 0$
  repeat
    $c \leftarrow c + 1$
    $t \leftarrow t_0 \cdot \gamma(c, C)$
    $S \leftarrow \{1, \ldots, p\}$
    until $S \neq \emptyset$ **do**
      Pick $s \in S$ and remove s from S
      $k_i \leftarrow s$
      $E^{k_i} := b^{k_i} - \mathbf{a}^{k_i} \cdot \mathbf{x}_i$
      $\delta_i := \frac{t}{t_0} \exp\left(\frac{-|E_i^{k_i}|}{t}\right) \mathbf{a}^{k_i}$
      if $(\mathbf{a}^{k_i}\mathbf{x}_i \leq b^{k_i} - \epsilon)$
        $\mathbf{x}_{i+1} := \mathbf{x}_i + \delta_i \mathbf{a}^{k_i}$
      else
        if $(\mathbf{a}^{k_i}\mathbf{x}_i \geq b^{k_i} + \epsilon)$
          $\mathbf{x}_{i+1} := \mathbf{x}_i - \delta_i \mathbf{a}^{k_i}$
      $i \leftarrow i + 1$
  until
    $c < C$
  Take $\mathbf{x}_{i+1}$ as an estimate of $\mathbf{x}_{max}$.
end

where t_0 depends on the average deviation from consistency (average inequality error) for the current solution $\mathbf{x}_i$ at the beginning of each cycle.

Intuitively, the behavior of the algorithm can be explained as follows. At high normalized temperature t/t_0, all equations with both high or low deviations from consistency lead to a significant correction of the current solution $\mathbf{x}_i$. Conversely, at low temperatures, only those equations with small deviations from consistency yield relevant corrections to the current solution. The convergence of the procedure is guaranteed because when t decreases to zero the amplitude of the modifications tends to zero. We refer the reader to [8, 10] for other motivations, justifications and more details about the overall greedy strategy as well as the above-mentioned procedure developed to find close-to-maximum consistent subsystems.

Although our simple PLIME algorithm is not guaranteed to lead to minimum size partitions, it turns out to be very effective experimentally. The very good results obtained for the two challenging applications we considered so far (optical flow segmentation [16] and time series state-space modeling [17]) suggest that it can be successfully applied to a wide range of problems.

The interesting analogies and differences between our PLIME algorithm and the Hough Transform are discussed in [8, 10].

4. CONCLUSIONS

A new combinatorial optimization approach is proposed for the domain decomposition and parameter estimation of piecewise linear models. The problem is formulated as that of finding a partition of the given inconsistent linear system into a minimum number of consistent subsystems (MIN PCS). A simple but efficient heuristic based on a greedy strategy and using variants of the perceptron algorithm is proposed for tackling MIN PCS. Our approach presents various advantages compared to classical alternatives such as regularization techniques, robust regression methods or the Hough transform. In particular, it does not suffer from limitations due to missing a priori knowledge on the domain decomposition or due to the absence of a dominant solution of the linear system. In contrast to robust regression techniques, it does not present any breakdown point. Moreover, since it compares very favorably with the above-mentioned techniques in terms of computational requirements, it is particularly suited to real-time applications.

References

[1] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, Vol. 76, No. 8, pp. 869–889, August 1988.

[2] J. Marroquin. Regularization theory and low-level vision. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 800–804, Cambridge, MA, 1995. MIT Press.

[3] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 4, pp. 413–424, July 1986.

[4] S. Haykin. *Adaptive filter theory*. Prentice Hall, Englewood Cliff, New Jersey, 1991.

[5] M. Basseville and I. Nikiforov. *Detection of abrupt changes*. Prentice Hall, Englewood Cliff, New Jersey, 1993.

[6] P. V. C. Hough. A method and means for recognizing complex pattern. Technical Report 3069634, U.S. Patent, 1962.

[7] D. H Ballard and C. M. Brown. *Computer vision*. Prentice Hall, Englewood Cliff, New Jersey, 1982.

[8] E. Amaldi and M. Mattavelli. A new combinatorial optimization approach for piecewise linear model estimation. *Submitted IEEE Trans. on Singal Processing*.

[9] P. Crescenzi and V. Kann. A compendium of np-optimization problems. *Available at: http:// www.nada.kth.se /~viggo /problemlist /compendium.html*, 1995.

[10] M. Mattavelli. *Motion analysis and estimation: from ill-posed discrete inverse linear problems to MPEG-2 coding*. PhD thesis, Swiss federal Institute of Technology, Lausanne, Department of Communication System, In preparation, 1996.

[11] E. Amaldi. *From finding maximum feasible subsystems of linear systems to feedforward neural network design*. PhD thesis, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[12] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, Vol. 147, pp. 181–210, 1995.

[13] M. L. Minsky and S. Papert. *Perceptrons: An introduction to computational Geometry*. MIT Press, Cambridge, MA, 1988. Expanded edition.

[14] M. Frean. A "thermal" perceptron learning rule. *Neural Computation*, Vol. 4, No. 6, pp. 946–957, 1992.

[15] E. Amaldi and C. G. Diderich. On the probabilistic and thermal perceptron training algorithms. Technical Report ORWP-8-94, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[16] M. Mattavelli and E. Amaldi. Using perceptron-like algorithms for the analysis and parameterization of object motion. In F. Girosi et al., editor, *Neural Network for Signal Processing V, Proceedings of the 1995 IEEE Workshop*, pages 303–312, Cambridge,Massachussetts, Aug. 31, Sept. 2 1995. IEEE Signal Processing Society.

[17] M. Mattavelli, J.M. Vesin, E. Amaldi, and R. Grueter. A new approach for piecewise linear modeling of time series. In *Proceedings of the 1996 Seventh IEEE Digital Signal Processing Workshop*, Loen, Norway, September 1996.