(k, m)-Segmentation Algorithm for Horizontal Lines

Lotan, Sagi* and Feldman, Dan*

*Robotic and big data lab,
University of haifa

Abstract. Division of an electrical signals into high and low, cuts a soccer video into filed of game and audiences shots... all of those are a thresholding problems. But when the noise get stronger the regular thresholding algorithm get weaker . In the km-segment mean problem we need to divide an input signal into a k-piecewise function when each piece is one of set M in size m.
we aproximate the solution for a constrained version of (k, m)-segmentation problem when the lines are horizontal . We can do that by trying every m size set in the input data as M and compute the k-segment of the signal while limit the segment to be from M set. then choose the one with minimum cost. What give us $cost < 4cost_opt$ by O(n chose m*) time using corset tocompute k-segment (fl14).
we show how our algorithm can divides better than a regular threshold and keep divide when we increase the noises even after threshold totally failed.

1 Introduction

1.1 Main Contribution

The main contributions of the paper are: (i) A solution for a constrained version of $(k, 2)$-segmantation problem(as given in **??**) when the lines are horizontal .(ii)And to test this new solution on a Infra Red drone's conroler in order to do the first step of decoding this data.

1.2 Problem Statement

The (k, m)-segment mean problem optimally fits a given discrete time signal of n points by a set of k linear segments over time, where $k \geq 1$ is a given integer. That is, we wish to partition the signal into k consecutive time intervals such that the points in each time interval are lying on a single line from one of an m types of lines, where $m \geq 1$ is a given integer to;

We make the following assumptions with respect to the data:(i) We assume the data is represented by a feature space that suitably represents its underlying structure; (ii) The content of the data includes at most k segments ,from at most m (for this paper $m = 2$) types. that we wish to detect auto- matically; An example for this are scenes of the audience and the field in a soccer video, a "bits" of an Infra Red signal (as we show in Section **??**), etc.
This motivates the following problem definition.

Definition 1 **((k, m)-segment mean).** *A set P in $\mathbb{R}^{d+1}$is a signal if $P = \{(1, p_1), (2, p_2), ..., (n, p_n)\}$ where $p_i \in \mathbb{R}^d$ is a point at time index i for evrey $i = [n] = \{1, ..., n\}$. For an integers $k \geq m \geq 1$, a $(k, m) -$ segmant is a k-picewise linear function :*

$$f : \mathbb{R} \to \mathbb{R}^d = \begin{cases} f_1'(x) & if\ x < d_1 \\ f_2'(x - d_1) & if\ d_1 x < d_2 \\ \cdots & \\ f_k'(x - d_{k-1}) & if\ d_{k-1} < x < d_k \end{cases}$$

where $\forall f_i : f_i' \in F$, and F is a group of lines from order m. this fanction f maps every time $i \in \mathbb{R}$ to a point $f(i)$ in $\mathbb{R}^d$.

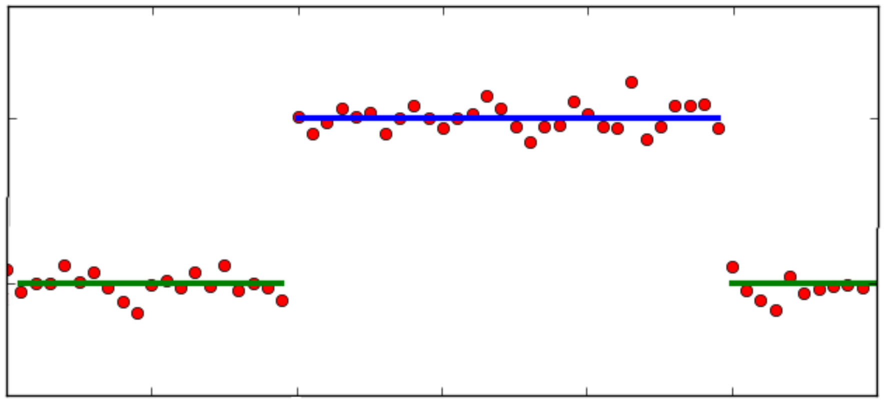


Fig. 1. This figure is an exemple to the $(3, 2)$-segment. the red dots are the input data and the lines are the segments when the blue is the low type and the grin is the high type.

The fitting error at time t is the squared distance between p_i and its corresponding projected point $f(i)$ on the (k, m)-segments. The fitting cost of f to P is the sum of these squared distances.

$$cost(P, f) = \sum_{i=1}^{n} \| p_i - f(i) \|_2^2 .$$

where $\| \cdot \|$ denotes the Euclidean distance. The function f is a (k, M)-segment mean of P if it minimizes $cost(P, f)$ and conducts the conditions above.

In this paper we interest in a constrained version of $(k, 2)$-segmantation ((k, m)-segmantation for $m = 2$) problem when the lines are horizontal.

2 The algorithm

If we lock at the problem we can see that (i) if we had the corct set F so the problem was esy sins all what we had to is to compute the k-segment(see Definition **??**) mean wehn eatch segmant is in F.(ii) And if we had the dividers of the segments on the (k, m)-segmat then each segment will be the mean height of the corresponding input points.

Definition 2 **(k-segment mean).** *The k-segment mean problem is a private case of the (k, m)-segment mean, when $k = m$.*

Definition 3 **(Bicriteria or (α, β)-approximation).** *For $\alpha, \beta > 0$, an $(\alpha, \beta) -$ approximation for the $(k, m) -$ segmentmean of P is a $(k) -$ segment g such that $cost(P, g) \leq \alpha * cost(P, f)$.*

Theorem 1 **(Bicriteria approximation [?]).** *Let $P = \{(1, p_1), ..., (k, p_k)\}$. Then an (α, β)-approximation for P can be computed in $O(dn)$ time where $\alpha = ..$ and $\beta = ..$*

For this solution to the km-segment problem we use i) Bicriteria. ii) BalancedPartition. As described in Feldman's 2014.

Algorithm 1 km - segmentation main

Input: A set $P = \{(1, p_1), ..., (n, p_n)\}$ in $\mathbb{R}^{d+1}$ and integer $k \geq 1$
Output: A 2-approximation to the (k, m)-segment mean of P
when $m = 2$.
Set $a \leftarrow \infty$
for $\{p_i, p_j \in P | 1 \leq i \leq j \leq n\}$ **do**
  Set $f \leftarrow \mathit{Bicriteria}(P, k, p_i, p_j)$
  $coreset = \mathit{BalancedPartition}(P, \varepsilon, cost(f), p_i, p_j)$
  $temp = \mathit{K\text{-}Segmentation}(coreset, k, p_i, p_j)$
  $best\_fit =$ Compute $temp$ such that $cost(P, temp)$ is minimized.
end for
return $best\_fit$

---

g_i is the line $y = p_i$

Algorithm 2 alternative for 1 - segment main

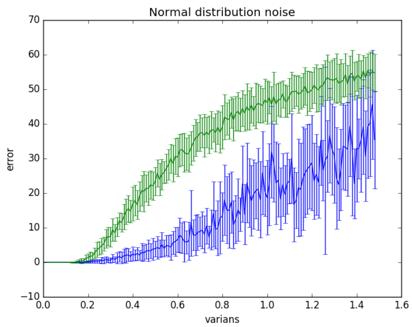
Input: A set $P \subseteq \mathbb{R}^{d+1}$ and two vectors $p_1, p_2 \in \mathbb{R}^d$
Output: An 2-approximation to the (k, m)-segment mean of P when $m = 2$.
if $cost(g_1, P) < cost(g_2, P))$ **then**
  return g_1
else
  return g_2
end if

Theorem 2 (solution cost). *This algorithm that we present will give a f' that $cost(P, f') \leq (2 + \varepsilon) * cost(P, f)$*

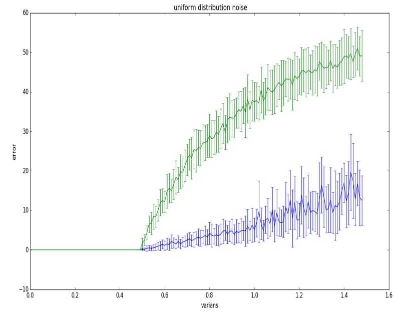
Proof. Without loss of generality if we lock on all of the dots on the segments that from the higher type . Let p_{opt} be the men of the corresponding of all those points, Let $cost_{opt}$ be the averege distance to p_{opt} and let p_s be the closest point to p_{opt}.Then the distanse $\Delta(p_{opt}, p_s)$ is smoler or equal to $cost_{opt}$ so for each point p in the data: $\Delta(p, p_s) <$

3 Experimental Results

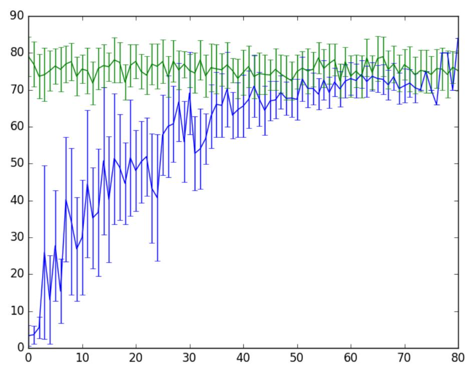
We now demonstrate the results of our algorithm on sintetic data in cuple of tests. We compare our algorithms against regular thresholding algorithem. we also show our algorithim preformens on a real life data token from an ir remote controlers.



(a) normal noise varians vs erorr



(b) uniform noise varians vs erorr



(c) number of segment vs erorr

Fig. 2. Figure 2a shows the division error as fanction of normal noise varians that added to the signal. For (k, m)-segmantation algorithim(blue) against threshold's(green). Figure 2b shows the same as 2a but for a uniform noise. Figure 2c shows the division error as fanction of the number of segments in the input signal (k, m)-segment(blue) against thresholding(green).

3.1 segmantation of data

We first examine the behavior of the algorithm on synthetic data which provides us with easy ground-truth, to evaluate the quality of the algorithem.We generate synthetic test data by drawing a discrete (k, m)-segment P whith $k = 15$ and $m = 2$, and then we add Gaussian and normal noise(one in eatch experament).

References

1. D . Feldman, G. Rosman, M. Volkov, J.W. Fisher III, and D. Rus Proc. *Coresets for k-Segmentation of Streaming Data*. 27th Conference on Neural Information Processing Systems (NIPS) 2014