Journal of Global Optimization (2006) 35: 103–109
DOI 10.1007/s10898-005-8465-2

On a Minimum Linear Classification Problem

BING LU[1], HONGWEI DU[2], XIAOHUA JIA[2], YINFENG XU[3] and BINHAI ZHU[4]
[1]*Department of Computer Science and Enginnering, University of Minnesota, Minneapolis, MN 55455, USA (E-mail: blu@cs.umn.edu)*
[2]*Department of Computer Science, City University of Hong Kong (E-mail: {hongwei,jia}@cs.cityu.edu.hk)*
[3]*School of Management, Xi'an Jiaotong University, China (E-mail: yfxu@xjtu.edu.cn)*
[4]*Department of Computer Science, Montana State University, Bozeman MT 59717, USA (E-mail: bhz@cs.montana.edu)*

(Received 6 June 2005; accepted in revised form 6 June 2005)

Abstract. We study the following linear classification problem in signal processing: Given a set B of n black point and a set W of m white points in the plane ($m = O(n)$), compute a minimum number of lines L such that in the arrangement of L each face contain points with the same color (i.e., either all black points or all white points). We call this the Minimum Linear Classification (MLC) problem. We prove that MLC is NP-complete by a reduction from the Minimum Line Fitting (MLF) problem; moreover, a C-approximation to MLC implies a C-approximation to the MLF problem. Nevertheless, we obtain an $O(\log n)$-factor algorithm for MLC and we also obtain an $O(\log Z)$-factor algorithm for MLC where Z is the minimum number of disjoint axis-parallel black/white rectangles covering B and W.

Key words: Approximation Algorithm, NP-complete, Sequence Detection, Signal Processing

1. Introduction

Intersymbol interference(ISI) arises in pulse-modulation systems due to limited bandwidth and is a primary impediment to reliable high-rate digital transmission or high-density data storage over narrow bandwidth channels, One approach to data detection for channels which suffer from ISI and additive Gaussian noise is to use a sequence of observation samples taken over certain number τ of future symbol intervals-decision delay in making a particular symbol decision. These types of detectors are called finite-delay sequence detectors. In [7, 12] Moon et al. implements such detectors using signal space partitioning. Their basic idea is that all possible finite-length signal sequences (the total number is $2^{\tau+1}$)are first mapped to a multi-dimensional vector space, then a systematic space partitioning method is proposed to divide the entire space into two distinct decision regions using a set of hyperplanes.

The discrete-time channel model is represented in [7] as $o_k = \sum_{i=0}^{L} f_i a_{k-i} + n_k$, where o_k is an observation sample, f_i represents the overall channel response ($f_0 \neq 0$) of a finite length L, a_k is the input symbol taken from $\{+1, -1\}$, and n_k is zero-mean additive white Gaussian noise. A sequence detector with a decision delay of τ makes a decision on the input symbol $a_{k-\tau}$ at time k based on observation samples $0_k(0 \leqslant k \leqslant \tau)$. Past decisions on the input symbols $a_{k-i}(i > \tau)$ are used to cancel ISI terms from observation samples under the assumption that the past decisions are correct. The observation samples available at the detector input after the cancellation are represented by $x_{k-j} = \sum_{i=0}^{\tau-j} f_i a_{k-j-i} + n_{k-j} = s_{k-j} + n_{k-j}$, where $0 \leqslant j \leqslant \tau, s_{k-j}$ is the noiseless signal. The detector finds a noiseless signal vector $\mathbf{s} = [s_k, \cdots, s_{k-\tau}]^T$ which maximizes the probability $p(\mathbf{x}|\mathbf{s})$ for a given observation sample vector $\mathbf{x} = [x_k, \cdots, x_{k-\tau}]^T$ and chooses the associated $a_{k-\tau}$ as the symbol decision. This decision process can be viewed as partitioning the $(\tau+1)$-dimensional observation space into appropriate non-overlapping decision regions. The corresponding decision boundary is piecewise linear, which can be represented by a set of hyperplanes.

Figure 1 shows the structure of the signal space detector. A finite number $\tau + 1$ of inputs to the linear discriminant functions are the channel output samples. Each linear discriminant function represents a hyperplane in $a(\tau + 1)$-dimensional signal space. Then a threshold detector determines which side of the corresponding hyperplane the observation vector $\mathbf{x}$ is located. At last, a Boolean logic function estimates the channel input symbol based on the location of the observation vector relative to each hyperplane. The main complexity of the detector is the number of hyperplanes. The goal is to minimize the number of hyperplanes for a given performance measure – the minimum distance between any signal and the decision boundary.

In [7], the detector design procedure starts with the opposite-class pairing of all signal subsets. Each pair defines a hyperplane separating them.

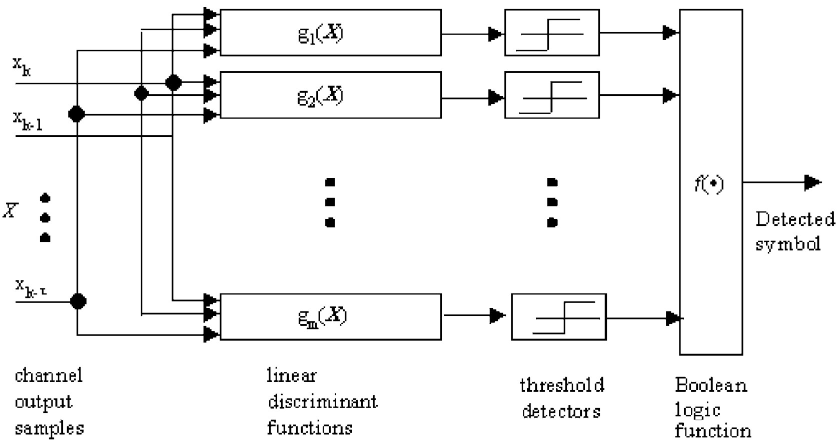

Figure 1. The structure of signal space detector.

By searching through these hyperplanes, a signal space detector with the minimal set of hyperplanes is obtained. It is shown that the proposed procedure also has a delay-constrained, asymptotically optimal detector if the performance measure $\beta_{\min}$ is set at half of the minimum distance $d_{\min}$ between any noiseless signals.

We can formulate the problem of finding minimum set of hyperplanes as: Given a set B of m black points, a set W of n white points in a space($m = O(n)$), a prescribed distance $\beta_{\min}$, find a minimum number of hyperplanes such that every pair of opposite-color points can be separated with distance at least $\beta_{\min}$, (i.e., the distance from the points to the hyperplane is at least $\beta_{\min}$).

In this paper we only focus on the two-dimensional (2D) version of this problem; more over, we assume that $\beta_{\min} = 0$. We call this restricted problem Minimum Linear Classification (MLC). We prove that even this restricted problem is NP-hard and we present an $O(\log n)$-factor approximation for this problem. We remark that a related problem of separating black/white points with minimum length tour (simple polygon) has been studied in [9] and another related problem of stabbing a set of objects (vertical segments or squares) with minimum number of lines has been studied in [4].

The paper is organized as follows. In Section 2, we prove the NP-hardness of MLC. In Section 3 we present two approximations for MLC. In Section 4, we close the paper with some open problems for further research in this topic.

2. Hardness Results

In this section, we establish the hardness result for MLC. We reduce the Minimum Line Fitting problem to it, which is a known NP-hard problem [10]. Given a set B of n (black) points with rational coordinates in the plane, the Minimum Line Fitting (MLF) problem is to compute a set of lines L^* with the minimum cardinality so that every point in B is on at least one line in L^* (or, equivalently, the lines in L^* fit all the points in B). Our result is as follows.

THEOREM 1. *The Minimum Line Fitting problem can be reduced to the Minimum Linear Classification problem in polynomial time.*

Proof. Given a set B of n black points, suppose that we want to find the minimum number of lines fitting B. We show below that MLF has a solution of size K if and only if an instance of MLC has a solution of size $2K$.

Let R be the set of all real numbers and let Q be the set of all rational numbers. Given an instance of MLF, we transfer each point $b = (x_b, y_b) \in B, x_b, y_b \in Q$ into a small triangle with three white points enclosing b. Let these three white points be $w_1^b = (x_1^b, y_1^b)$, $w_2^b = (x_2^b, y_2^b)$ and $w_3^b = (x_3^b, y_3^b)$, where all the coordinates of $w_i^b, i = 1, 2, 3$ are in R. For any two points $b, d \in B$ we can choose the coordinates of w_i^b and $w_i^d, i, j = 1, 2, 3$, such that there exist two *white lines* defined by w_i^b and $w_j^d, j = 1, 2, 3$, whose slopes bound that of line (b, d). Furthermore, we can choose $w_i^b = 1, 2, 3$, for all b, such that no three white points are collinear and no white point is collinear with any *black line* (i.e., a line defined by two black points). As there are only a finite $(O(n^2))$ possible black lines defined by points in B we can construct all these white points in polynomial $(O(n^2))$ time. We loosely call this procedure *perturbation.*

We are now ready to make some claims. First of all, notice that given w_1^b, w_2^b, W_3^b and b we must use at least two lines to separate b from the three white points. Secondly, given any two such white triangles, each enclosing a black point, we must use at least two lines to separate the two black points from the six white points. Thirdly, given any k black points in B, if we could use only one line to fit these k black points then we can use two lines to separate the k black points from those corresponding $3k$ white points.

Consequently, MLF has a solution of size K if and only if the minimum linear classification problem has a solution of size $2K$. This reduction clearly takes polynomial time. □

From the above theorem, MLF is NP-hard. As it is easy to see that MLC is in NP, we have the following corollaries.

COROLLARY 1. *Minimum Linear Classification is NP-complete.*

Notice that for a (minimization) optimization problem Π, we say that an approximation algorithm A achieves an approximation factor of ρ if for every instance of Π the solution value returned by A is at most ρ times the corresponding optimal solution value. We also say that A is a p-approximation of Π. As both MLC and MLF are NP-complete, it would be interesting to approximate them. Theorem 1 implies the following corollary regarding efficient approximations for MLC and MLF.

COROLLARY 2. *If there is a C-approximation for Minimum Linear Classification then there is a C-approximation for the Minimum Line Fitting problem.*

Proof. For the instance we use in the proof of Theorem 1, it is clearly that if the optimal solution of MLF has size O_{MLF} then the optimal solution for MLC must have size at least 2. O_{MLF}. If there is a

C-approximation for MLC then we can run the approximation algorithm on the instance we use in the proof of Theorem 1 and obviously that produces a C-approximation for the MLF problem. □

The best approximation for MLF has a factor $2\log_{n+1}$ – just formulate it as a set-cover problem where each line is viewed as a set and the points on it are its elements. Although the general set-cover problem cannot be approximated with a factor better than $\Omega(\log n)$ [13], we are not able to make the same claim for MLF and it is not known whether a constant-factor approximation for MLF exists or not. Corollary 2 shows that approximating MLC is at least as hard as approximating MLF. In the next section we present two approximation algorithms for MLC.

3. Approximation Algorithms

In this section, we present two approximation algorithms for MLC. The first approximation is to formulate it directly as a set-cover problem as follows. Connecting all black points and white points, we have mn segments whose endpoints have different color. We call these set S.

It is clear that our problem is basically to find the minimum number of lines stabbing all the segments in S. In fact, we need only to search a set $L(S)$ of $O((m+n)^2)$ candidate lines for MLC: through each segment $(p, q), p, q \in B \cup W$ we rotate the line (p, q) around the center of the segment $(p, q), o_{pq}$, in clockwise order until it hits another point r in $B \cup W$ – during this process all the lines are topologically equivalent, i.e., separate the same subsets of points in $B \cup W$. We take the line l_1 bisecting $\angle po_{pq}r$ and put it into $L(S)$. If we rotate line (p, q) around o_{pq} in counterclockwise order we obtain another line 1_2 for $L(S)$. So each pair $p, q \in B \cup W$ contributes two lines for $L(S)$. Now the problem is easy, we have a set-cover problem which we want to use the minimum number of lines in $L(S)$ to stab all the segments in S whose endpoints have different colors. Each line in $L(S)$ is viewed as a subset which consists of the segments in S it stabs.

For set-cover, there is a standard greedy algorithm (in our case, each time we choose the line in $L(S)$ which stabs the maximum number of segments in S). With the standard analysis, the greedy algorithm achieves an approximation factor of $\log \frac{n(n-1)}{2} + 1 \leqslant 2\log n + 1$ [1, 6, 8]. Interested readers can also refer to some textbooks for simplified analysis [2]. We thus have the following result.

COROLLARY 3. *There exists a* $(2\log n + 1)$*-approximation for Minimum Linear Classification.*

In the following, we present another algorithm with an approximation factor of $2\log Z + 1$ where Z is the minimum number of disjoin axis-parallel boxes containing only white (black) points. Our motivation is that even though Z could be quadratic, in practice it would be relatively small. The algorithm works as follows.

Input: A set B of n black points, a set W of m white points.

Step 1. Draw horizontal lines through midpoints of the x-intervals defined by all of the points in $B \cup W$ sorted along x-axis. Let this set of lines be H.

Step 2. Draw vertical lines through midpoints of the y-intervals defined by all of the points in $B \cup W$ sorted along y-axis. Let this set of lines be V.

Step 3. Let the arrangement of $H \cup V$ be $A(H \cup V)$. Color each cell of the arrangement black (white) if it contains a black (white) point.

Step 4. Compute the maximum connected components of black (white) cells. We have a set of black rectilinear polygons possibly with white or empty holes and symmetrically, we have a set of white rectilinear polygon possibly with black or empty holes.

Step 5. Use Imai and Asano′s algorithm [5] to decompose each of the rectilinear polygon (with or without holes) into the minimum number of rectangles. Let the set of all black (white) rectangles be $R_B(R_w)$.

Step 6. For each rectangle in $R_B \cup R_W$, compute convex hull for all the points inside it. Let Z be the number of convex polygons obtained. Now use the greedy algorithm to solve the following set-cover problem: the elements in this set system are the shortest segments connecting two convex polygons with different colors and the subsets are the (slightly perturbed) inner tangents of these Z^2 black/white polygon pairs.

The algorithm runs in $O(n^2)$ time as we must compute the arrangement of $O(n)$ lines. Although in the worst case, $Z = \Omega(n^2)$ (this happens when we have $\Omega(n^2)$ connected components in Step 4), in practical situation it should be much smaller. Summarizing the analysis, we have

THEOREM 2. *The Minimum Linear Classification problem can be approximated with a factor of* $2\log Z + 1$, *where* Z *is the minimum number of disjoint axis-parallel black/white rectangles covering* B *and* W.

4. Remarks

In this note we show that the Minimum Linear Classification problem is NP-complete. Our reduction shows that if there is an $o(\log n)$-factor approximation for it then there is also an $o(\log n)$-factor approximation for the MLF problem. Therefore, it is interesting to know whether there is an $o(\log n)$-factor approximation for the MLF problem and whether the ratio $\log n$ is the lower bound for approximating the MLC problem. The more interesting problem is to investigate the original problem in high dimensions for which nothing is known so far.

References

1. Chvátal, V. (1979), A greedy heuristic for the set-covering problem, *Mathematics of Operations Research* 4, 233–235.
2. Cormen, T. Leiserson, C. and Rivest, R. (1990), *Introduction to Algorithms*, The MIT Press.
3. Forney, G. D. Jr., (1972), Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference, *IEEE Transactions Information* 18(30), 363–378.
4. Hassin, R. and Megiddo, N. (1990), Approximation algorithms for hitting objects with straight lines, *Discrete Applied Mathematics* 30, 29–42.
5. Imai, H. and Asano, Ta. (1986), Efficient algorithms for geometric graph searching problems, *SIAM Journal on Computing* 15, 478–494.
6. Johnson, D. (1974), Approximation algorithms for combinatorial problems, *Journal of Computer and System Sciences* 9, 256–278.
7. Kim, Y. and Moon, J. (2000), Multi-dimensional signal space partitioning using a minimal set of hyperplanes for detecting ISI-corrupted symbols, *IEEE Transactions on Communications*, April 2000.
8. Lovász, L. (1975), On the ratio of optimal integral and fractional covers, *Discrete Mathematics* 13, 383–390.
9. Mata, C. and Mitchell, J. (1995), Approximation algorithms for geometric tour and network design problems, *Proceedings of the 11th ACM Symposium on Computing Geometry*, 360–369.
10. Megiddo, N. and Tamir, A. (1982), On the complexity of locating linear facilities in the plane, *Operations Research Letters* 1, 194–197.
11. Moon, J. and Carley, L. R. (1990), Performance comparison of detection methods in magnetic recording, *IEEE Transactions on Magnetics* 26(6), 3155–3172.
12. Moon, J. and Jeon, T. (1998), Sequence detection for binary ISI channels using signal space partitioning, *IEEE Transactions on Communications* 46(7), 891–901.
13. Raz, R. and Safra, S. (1997), A sub-constant error-probability low-degree test, and a subconstant error-probability PCP characterization of NP, *Proceedings of the 29th ACM Symposium on Theory of Computing (STOC97)*, 475–484.