A Linear Time Algorithm for Partitioning a set of points in the Plane*

Y. T. Ching and D. T. Lee
Institute of Information Science
Academia Sinica
Taipei, Taiwan
Republic of China

March, 1985

I Introduction

This note describes a linear time algorithm for finding two lines, L_1 and L_2, in the plane such that they divide a set S of n=4m, m 0, points into four equal subsets, i.e., exactly one fourth of the points lie in one quadrant. It turns out that this problem is not easier even if one of these two lines is given. In the following, we shall assume L_1 is given. It is not difficult to see that we can obtain L_1 by choosing a vertical line passing through the median of x-values of the points or a line passing through the median of the polar angles of the points with respect to a fixed point outside of the convex hull of the set of points.

The algorithm described here is based on the geometric duality transform between points and lines in the plane through a unit circle. A point P in the plane with (a,b) as coordinates is said to be the dual of the line lP : ax + by + 1 = 0 and vise versa[CGL]. Assume without loss of generality that the origin, O, of the unit circle is external to the convex hull of the set of the points. To simplify the problem, we let one of the lines, say L_1 which divides the points into two equal halves, pass through O. We shall search for a line L_2, which divides

* This work was done while the aurthors are on leave from Dept. of EECS, Northwestern University, Evanston, IL 60201.

those two sets of points on both sides of L_1 into two equal subsets respectively (Fig-1).

We transform the points of S into a set IS of lines in the dual plane and let L_1 be the Y-axis and O be the origin of our new XY-coordinate system in the dual plane. Since L_1 divides the points into two equal subsets, half of the lines in the dual plane have negative slopes and the other half have positive slopes with respect to the new coordinate system.

Let the image of L_2 be IL_2. The system has the following properties : The images of the points to the right of line L_1 have negative slopes. A point in this set which is above or below L_2 has the image intersecting the directed line $\overrightarrow{OIL_2}$ below or above IL_2 respectively. The same thing happens to the points to the left of line L_1 except that they have positive slopes. Given a line L_3 which has slope 0 and passes through IL_2. The line L_3 and the line OIL_2 passing through O and IL_2 constitute a new coordinate system and the resulting four quadrants are denoted Q_i, $i=1,2,3,4$(Fig-2). Let QS_i, $i=1,2,3,4$, denote the set of lines in IS which do not pass through the quadrant Q_i. The set IS of lines in the dual plane is called '<u>balanced</u>' with respect to L_3 and $\overleftrightarrow{OIL_2}$ if $|QS_i| = n/4$, $i=1,2,3,4$. Thus the problem of partitioning a set S of points in the plane is equivalent to the problem of finding the coordinate system for the set IS of lines such that IS is balanced with respect to the coordinate system.

Let the set of points to the left (right) of L_1 be denoted S_L (S_R) and the set of images of S_L (S_R) be denoted IS_L (IS_R). Let P be a point in the dual plane. The point Pt is a '<u>center</u>' for IS_L and IS_R, if for all lines passing through P, there are at least one fourth of the intersection points with IS_L and IS_R respectively that are on one side of the line.

The median chain for the set IS_L of lines is defined as the following : Draw a line L passing through O. Among the intersections of L with the lines in IS_L, we say P_i precedes P_j if (1) $Y_i>0$ and $Y_j<0$, (2) $Y_i<Y_j$ if $Y_i*Y_j>0$, (3) $X_i<X_j$ if $Y_i=Y_j=0$. Fig-3 illustrates the ordering of the points. Given an ordered list of the intersection points, we have a median. The median chain is the union of the medians of the ordered lists obtained by rotating the line with pivot at O. The p'th chain of IS_L is defined in a similar manner except that it is the union of p'th intersection points. Let $F_i(X) = a_iX + c_i$, $i=1,2, \ldots ,n$, be a set of linear functions. The maximum <u>envelope</u> is the union of $\max(F_i(X))$ for all x. Thus the origin of the unit circle is above the maximum envelope of the set of lines. Therefore if we sweep a vertical line from left to right and define the ordering of the intersections in their y-values, we can obtain the same median chain as obtained by rotating a line with pivot at O. For simplicity of computation, we shall use a vertical line in subsequent procedures. Since the line in IS_L (IS_R) has positive (negative) slope, if we sweep a line from left to right, the median chain of IS_L (IS_R) always goes upward (downward). The intersection Pt of the median chain of IS_L and IS_R can be used as the <u>center</u>. Since the set IS of lines is balanced with respective to the coordinate system with Pt as the origin and the horizontal line passing through Pt and the line OPt as the coordinate axes. Therefore, from now on, we shall look for the intersection of the two median chains.

Lemma 1 : There is exactly one intersection of the two median chain in the dual plane.

Proof From left to right, one chain goes upward and the other goes downward, so these two chains must intersect. Since these two sets of lines have disjoint slopes, they can intersect only once (Fig-4). The lemma follows.

II The Algorithm

We give a general algorithm that finds the intersection of the p'th chain and q'th chain of two sets of lines having disjoint slopes without computing the whole chains. We shall use prune-and-search approach[LP,M] in which we drop a fraction α of n lines in each iteration in O(n) time so that the total run time is O(n). That is, if T(n) denotes the time require for this task, and have $T(n) = T((1-\alpha)n) + O(n)$, for some $0<\alpha<1$, which is O(n)[AHU,M]. For this problem, we are searching for the intersection of two median chains. Thus, in the first iteration, we let p and q denote the median chains of the two subsets of lines. Subsequently, since we have dropped several lines in each iteration, we shall solve the subproblem of finding the intersection of the p'th chain and q'th chain of two subsets of lines instead of the intersection of two median chains. We describe the first iteration only, since the remaining iterations are similar.

Let us make the following observation. Draw a vertical line to the left of the center. The intersection of the vertical line with the median chain of IS_R is higher than that of IS_L(Fig-5). Note that the intersection of a vertical line with the median chain of IS_L (IS_R) can be obtained by selecting the median of the intersections of the vertical line and the lines in IS_L (IS_R). That is, we may answer the location of the center being to the left (right) of a given vertical line in linear time. We shall work on the line in IS_L first. To select the vertical line LV, we divide IS_L into two subsets by the median slope and form pairs of lines from these two subsets, one line from the set with larger (than the median) slopes of lines and the other is from the set with smaller slopes. Let LV be a line passing through the median (in X-value) of the intersections of the pairs of lines(Fig-6). Suppose that the center is to the right of LV. We now work on the pairs of lines whose intersections are to the left of LV. Let MX denote this subset of lines. Note that, the number of lines in IS_L is n/2, so the number of lines in MX is n/4.

We shall search for a line LM intersecting LV so that by determining on which side of LM the center lies, we can drop a fraction of lines. Among the intersections of LV with the lines in MX, let MP be the median (in Y-value) of them. We draw a line LM with the median slope of IS_L and passing through MP.

Lemma 2 : There are at least one fourth of the lines in MX (n/16) that are totally above or below LM in the half-plane to the right of LV.

Proof Since MP is the median of the intersection points, for a pair of lines not totally separated by LM in the half-plane to the right of LV, there exists another pair of lines on the other side of LM (Fig-7).

If the center is above LM, i.e., the center lies in the right upper quadrant, then the center does not lie on the lines totally below

LM(Fig-8). So we can drop those lines because they will not affect the answer. Note that the number of lines that can be dropped is at least n/16.

The algorithm that determines on which side of LM the center lies is based on the following observation.

The slope of LM is the median of IS_L so it must be positive. The slopes of the line segments on the median chain of IS_R are negative. LM and the median chain of IS_R can intersect at one point. Let Pt be the intersection of LM and the median chain of IS_R. Point Pt divides the chain into two subchains. There is only one center which must lie on one of the subchains so that by locating the ceneter with respect to a vertical line passing through Pt, we can answer the location of the center with respect to LM (Fig-9). It is easy to see that this process takes $O(n)$ time.

We shall apply the procedures stated above to IS_R. After this iteration, we can drop at least n/16 lines from IS_L and IS_R respectively. The number of lines that have been dropped is n/8. Depending on how the lines are dropped in the first iteration, we update p and q accordingly in the next iteration. That is, if we have dropped r lines, $r \geq n/16$. The number p in the next iteration is $n/4 - r$.

<u>Theorem</u> : The center for two sets of lines with disjoint slopes can be obtained in $O(n)$.

As a result of the theorem, we have the following corollary.

Corollary : Partitioning a set of points in the plane into four subsets of the same size can be solved in O(n).

To summarize the algorithm, the pseudo code is as the following:

Program partition (S, L_1, L_2);
    /* S is a set of points */
    /* L_1, L_2 are two lines that divide the set of points into four equal subsets */

Procedure test_center (LN,LD);
    /* LN is a set of lines which will be dealed with */
    /* LD is a set of lines which will be dropped */
begin
  select the median slope of LN;
  form pairs for the set LN;
  solve the intersection points for all pairs;
  search LV for LN;
  determine the location of the center to LV;
  solve all the intersections of LV with the surviving lines;
  search MP and LM;
  determine the location of the center to LM;
  LD := the lines in LN which can be dropped;
end;

Procedure center (IS_L, IS_R, p, q, CTR);
    /* ISD_L is a set of lines which can be dropped */
    /* CTR is the center in the dual plane */
begin
  test_center (IS_L, ISD_L);
  test_center (IS_R, ISD_R);
  $IS_L := IS_L - ISD_L$;
  $IS_R := IS_R - ISD_R$;
  if $(IS_L \;\; 2)$ or $(IS_R \;\; 2)$

```
        then CTR := the answer
        else
            begin
               update p,q accordingly;
               center (IS_L,IS_R,p,q);
            end;
end;

begin
   obtain L_1;
   transform S to L; /* L is a set of lines in the dual plane */
   divide L into IS_L and IS_R;
   p := median;
   q := median;
   center (IS_L,IS_R,p,q);
   L_2 := the dual of CTR:
end.
```

Applications

1. An O(nlogn) time algorithm to create the quad-tree.

Quad-tree is a data structure for half-planar range query. The algorithm stated above takes O(n) for each iteration. For the second iteration, we shall solve four subproblems each with one fourth of the original size. That is, $T(n) = 4*T(n/4) + O(n)$. The total complexity is O(nlogn).

2. Linear time algorithm for the 'center' of a set of points in the plane.

The 'center' here is defined as a point in the plane such that for all lines passing through this point, at least one fourth of the points are on one side of this line[Y].

In [Y], it takes $O(n^2)$ to locate the 'center'. As a result of the corollary, we have an $O(n)$ optimal algorithm.

References :

[AHU] Aho, A. V., J. E. Hopcroft and D. Ullman, "The design and analysis of Computer Algorithms," Addison-Wesley.

[CGL] Chazelle, B. M., L. J. Guibas and D. T. Lee, "The power of geometric duality," Proc. IEEE 24th symp. Foundation of Comput. Sci. Nov. 1983, 217-225.

[LP] Lee, D. T. and F. P. Preparta, "Computational Geometry - A Survey," IEEE Trans. on Computer Vol. C-33, No. 12, Dec. 1984, 1072-1101.

[M] Megiddo, N. "Linear-time algorithms for linear programming in R^3 and related problems," SIAM J. Comput. Vol. 12, No. 4, Nov. 1983, 759-776.

[Y] Yao, F. F., "A 3-space partition and its applications," Proc. 15th. Annual Symp. Theory of Computing, Apr. 1983, 258-263.

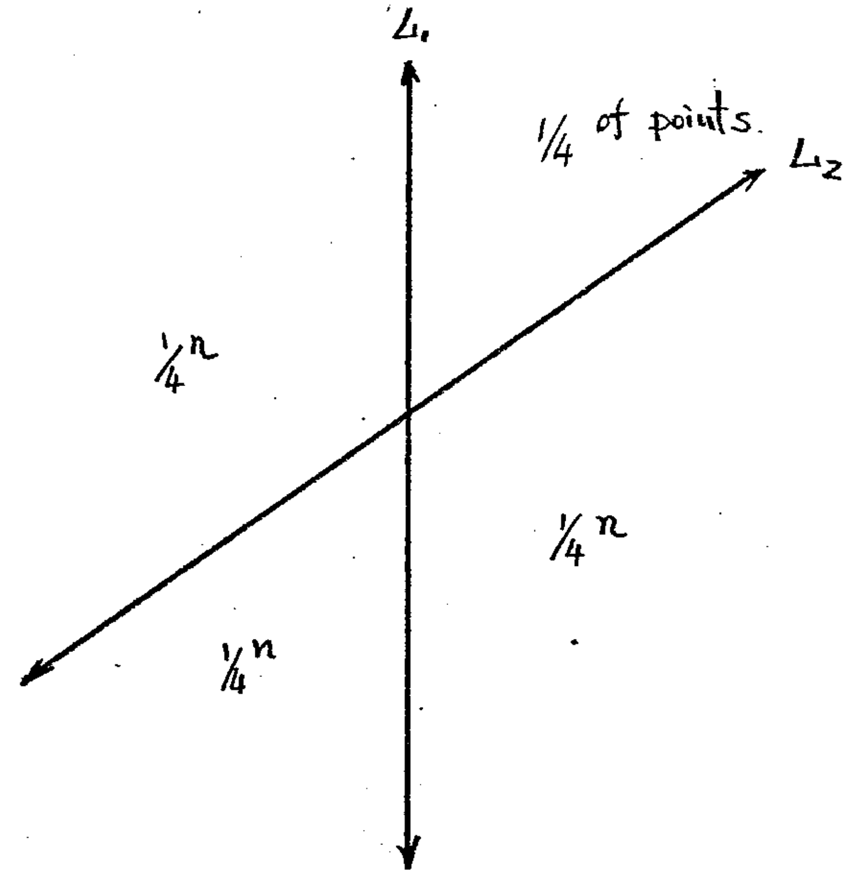


Fig - 1. L_1 and L_2 divide n points into four subsets.

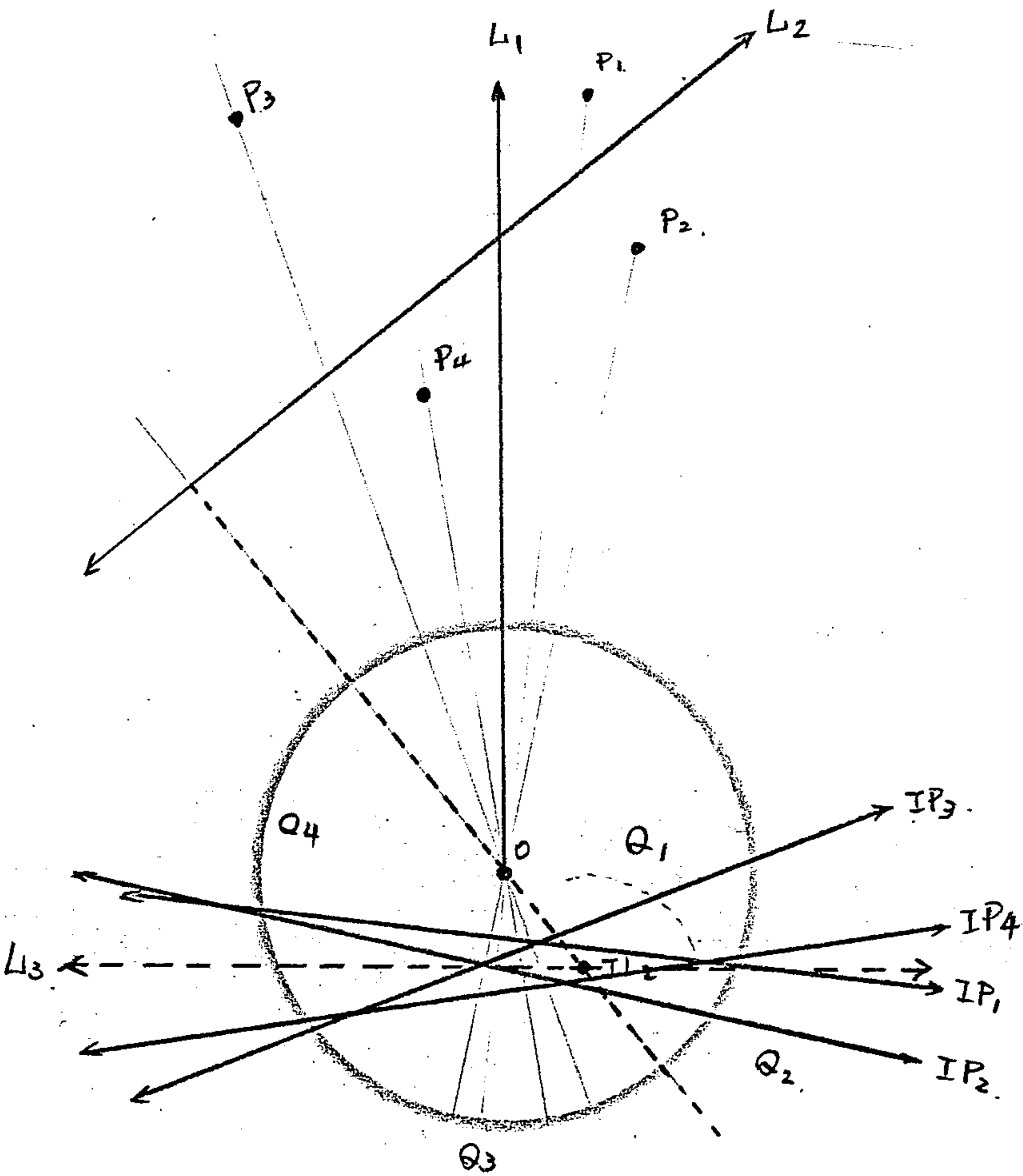


Fig 2: $IP_2 \in Q_1$, $IP_3 \in Q_2$, $IP_4 \in Q_4$

$IP_1 \in Q_3$.

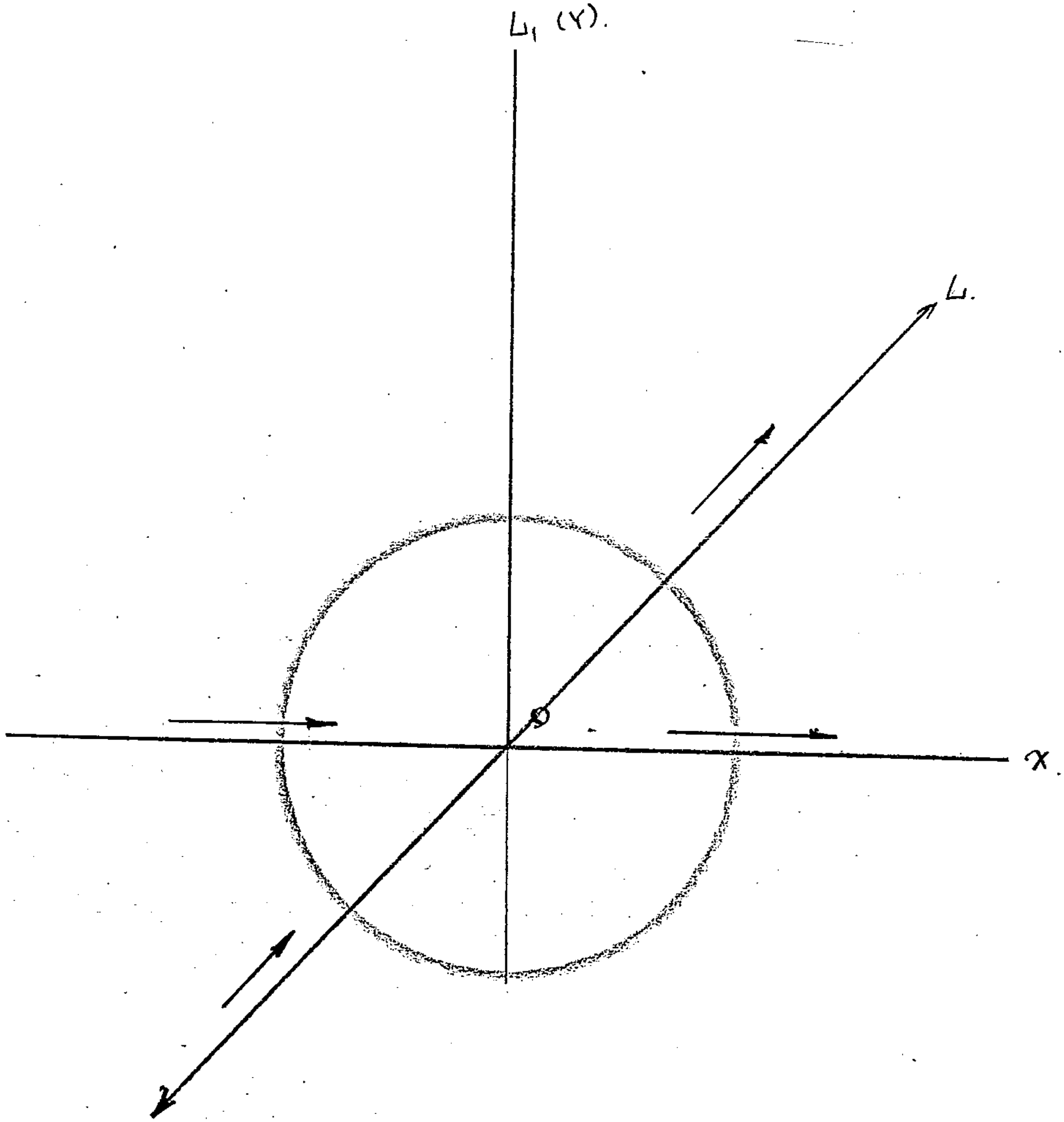
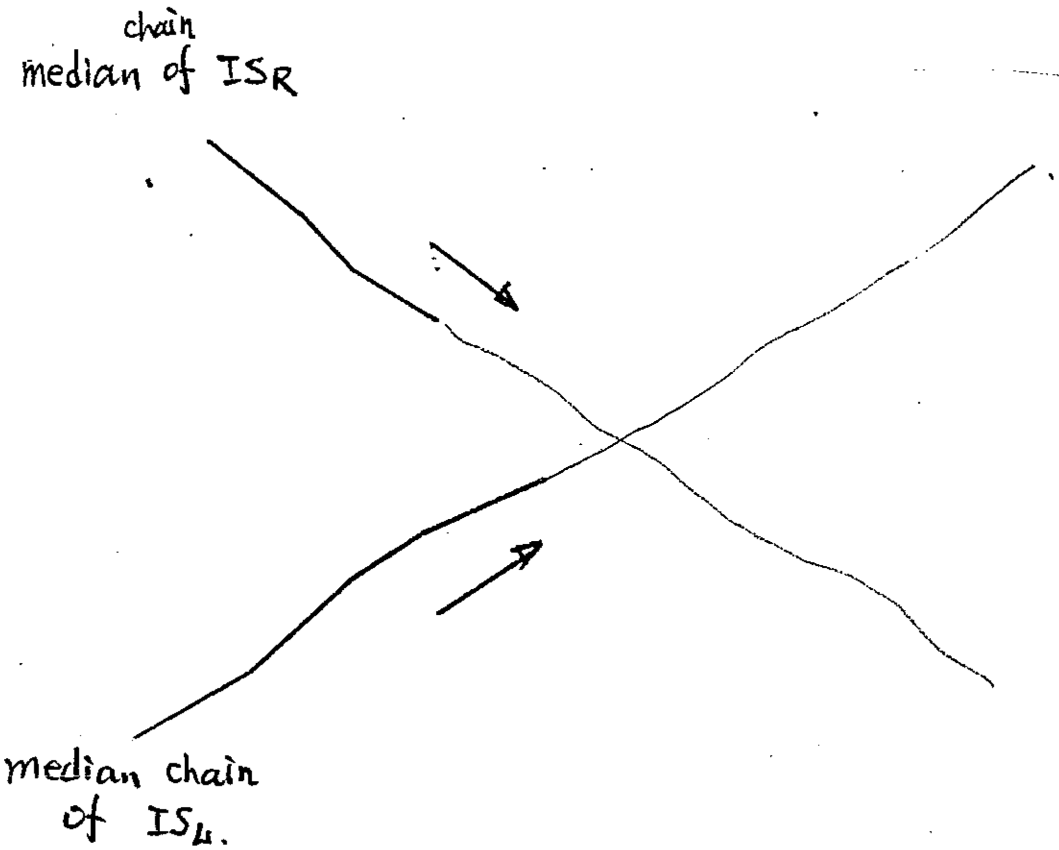
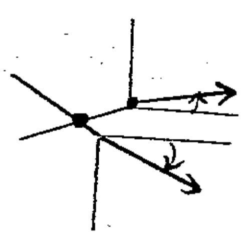


Fig 3 : The ordering of points.



a) they must intersect.



b) once intersect, they will never meet again.

Fig 4 : Illustration of Lemma 1.

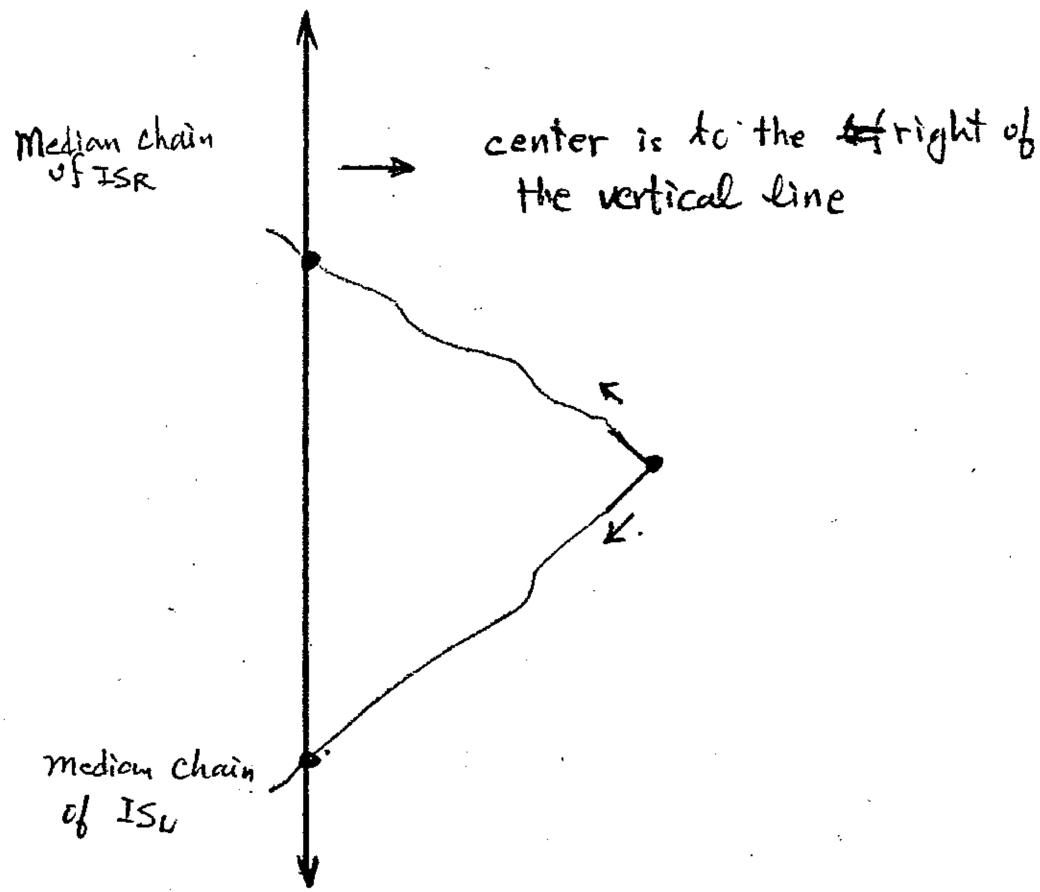

Median chain of ISR
center is to the ~~le~~ right of the vertical line
median chain of ISL

Fig 5

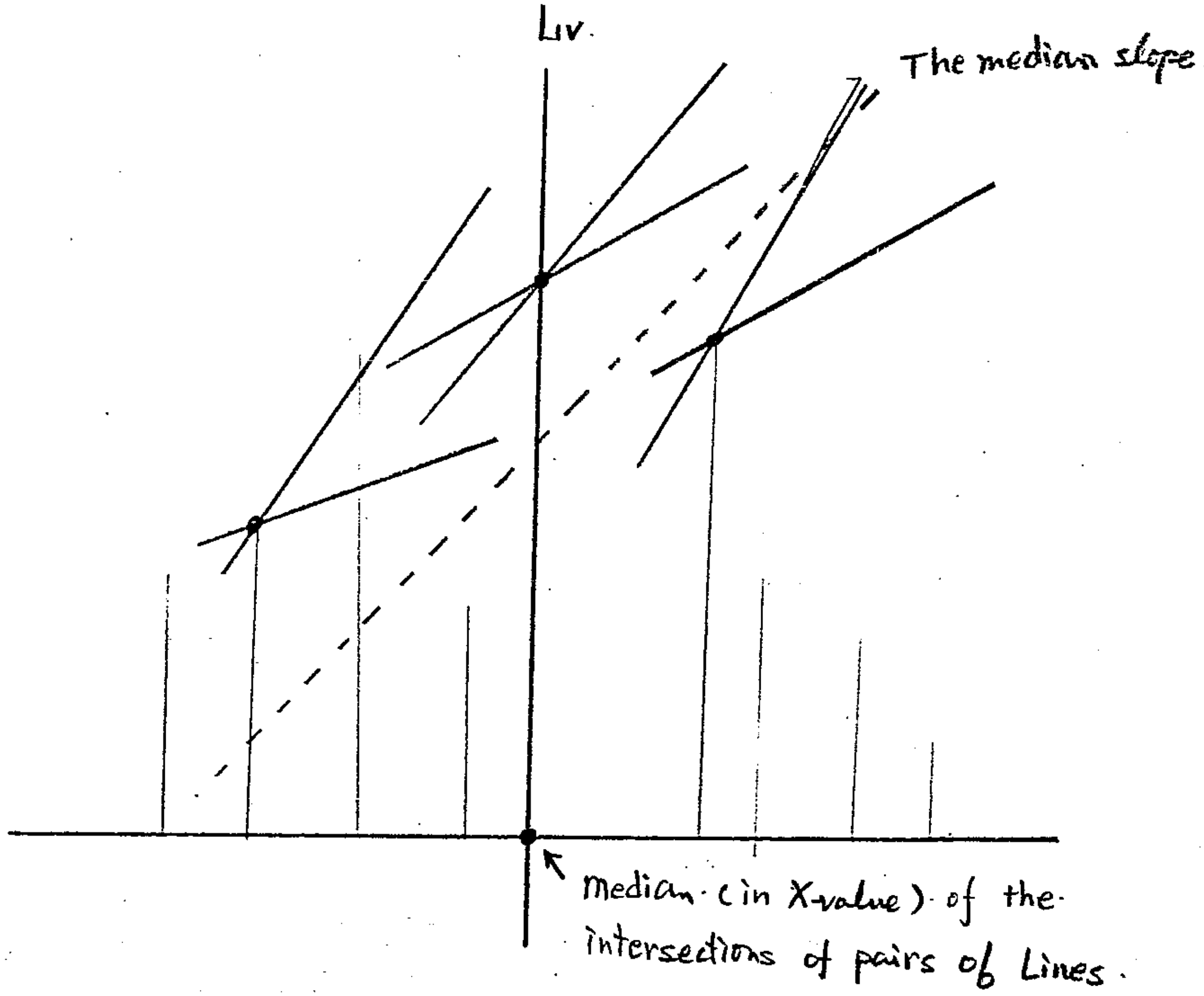

Liv.
The median slope.
median (in X-value) of the intersections of pairs of Lines.

Fig 6.

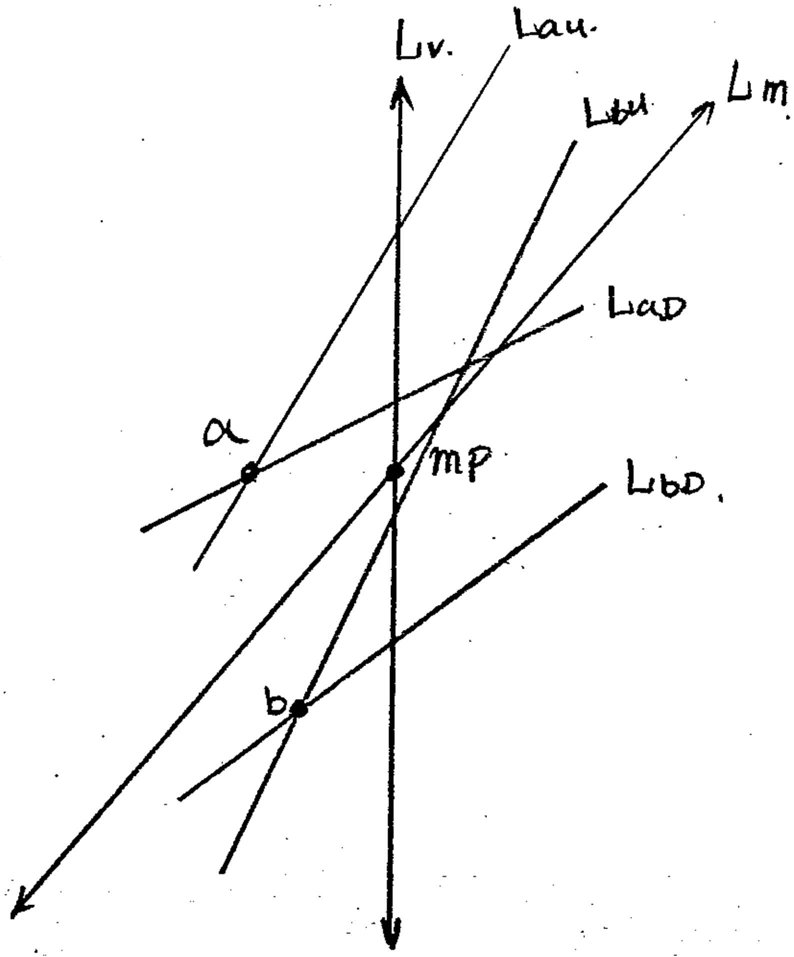


Fig 7: For a pair a, LaD is not totally above Lm.
there is a pair b, s.t. Lbu is not totally ~~above~~ below Lm (Since MP is the median)

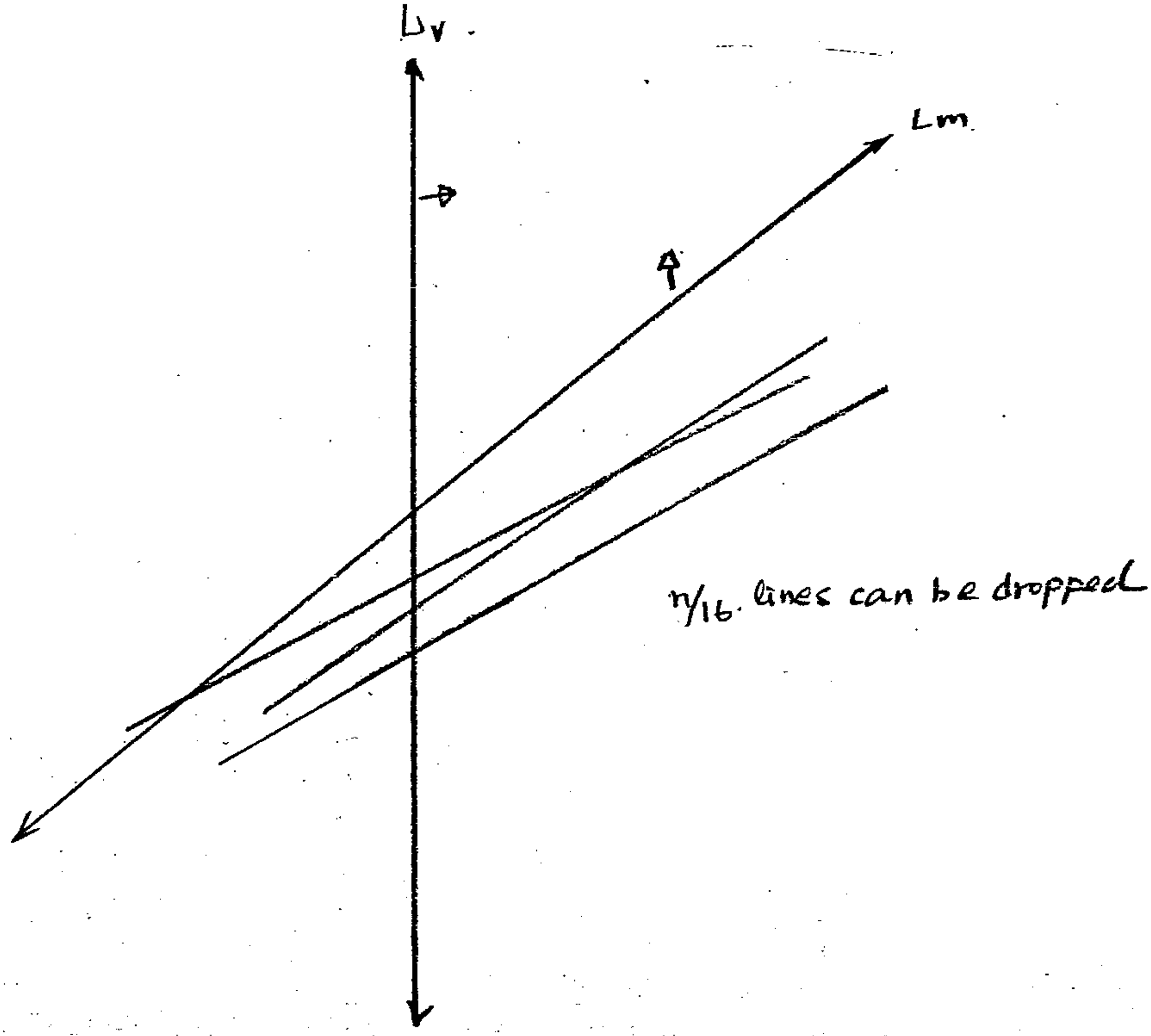


Fig 8.

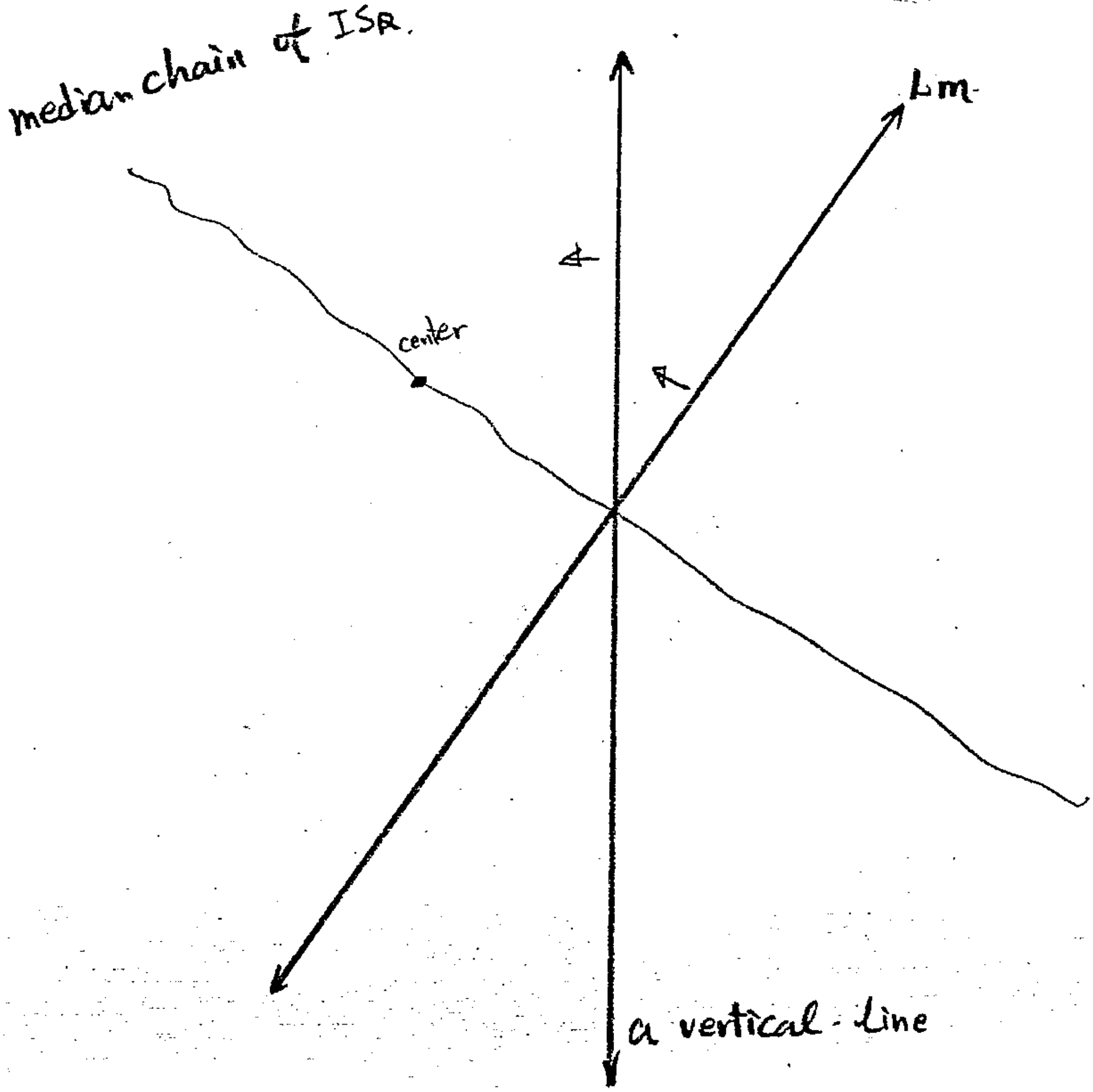


Fig 9. center is to the left of the vertical line
⇒ center is above Lm.