F. P. Palermo

A Network Minimization Problem

This letter deals with a problem which arose during a study of communications networks. The study has led to the mathematical problem described here.

A communications network consists of a set of stations at various geographical locations. Each station may be connected, through a switching center, to other stations in the system. The problem is to determine which stations are to be connected on the same line and where the switching center is to be located so that the cost of the lines is minimized; at the same time, a specified level of service is provided. The cost of the lines is assumed to be proportional to their lengths. This assumption is valid if the lines are sufficiently long (depending on the tariff rates). One is given the message load at each station and the level of service is specified by saying that the lines are to be loaded to a percentage of their capacity.

This has led to the following general problem:

> *Find the stations which are to be connected together in one line and then connect them into a center so as to minimize the total cost of the network.*

In this letter we will consider a simplified version of this problem in which we assume that a set of stations has been connected together on a line and that these lines are to be connected into the center. In the general problem the grouping is undetermined, and we are looking for an optimal grouping as well as for an optimal tying of these groups to the center. In the simplified version, the grouping is predetermined and we can treat each group as if it were an individual station. This form of the problem is identical to that of a relatively small number of stations being connected individually to a center.

Analytic method

Let us assume that each station is to be connected directly to the switching center using m_i lines, and that the stations lie in a plane. This problem can be formulated mathematically as follows: Given a set of N points, P_i, in a plane and a set of weights m_i, a point C is to be found so that the expression $\sum_{i=1}^{N} m_i l_i$ is minimized (where l_i is the distance from P_i to C).

Introducing a coordinate system in the plane, one can express l_i as a function of the coordinates (x, y) of C by the formula $l_i(x,y)=\sqrt{(x-x_i)^2+(y-y_i)^2}$, where (x_i, y_i) are the coordinates of the point P_i.

The problem thus reduces to the minimization of the function F, defined by the formula

$$F(x,y)=\sum_{i=1}^{N} m_i l_i(x,y)\,.$$

We shall prove that the surface defined by F, namely those points (x, y, z) such that $z=F(x,y)$, has a unique minimum if and only if the set of points P_i are not collinear, i.e., do not all lie on the same line.

To prove that the function F has exactly one minimum, we establish four lemmas.

In the first lemma, we compute the discriminant $F_{xx}F_{yy}-F_{xy}^2$ and in Lemma 2 observe that it is always positive if the points P_i are not collinear. Using this result, we then establish Lemmas 3 and 4 so that the proof of the main theorems will follow easily. We now proceed with the formal proof:

Lemma 1. $F_{xx}(x,y)F_{yy}(x,y)-F_{xy}(x,y)^2=\sum_{i<j}\dfrac{m_i m_j}{(l_i l_j)^3}$ $\times\{(x-x_i)(y-y_j)-(x-x_j)(y-y_i)\}^2$, *where* F_{xx}, F_{yy} *and* F_{xy} *are the second partial derivatives of* F *with respect to* x *and* y.

Proof: A straightforward computation gives:

$$F_{xx}(x,y)=\sum_i \frac{m_i(y-y_i)^2}{l_i^3}$$

$$F_{yy}(x,y)=\sum_i \frac{m_i(x-x_i)^2}{l_i^3}$$

$$F_{xy}(x,y)=\sum_{i=1}^{N} \frac{m_i(x-x_i)(y-y_i)}{l_i^3}\,.$$

Then

$$F_{xx}F_{yy}-F_{xy}{}^2=\sum_{i,j=1}^{N}\frac{m_im_j(y-y_i)^2(x-x_j)^2}{l_i^3\,l_j^3}$$
$$-\sum_{i,j=1}^{N}\frac{m_im_i(x-x_i)(x-x_j)(y-y_i)(y-y_j)}{l_i^3\,l_j^3}$$
$$=\sum_{i,j=1}^{N}\frac{m_im_j}{l_i^3\,l_j^3}[(y-y_i)^2(x-x_j)^2$$
$$-(x-x_i)(x-x_j)(y-y_i)(y-y_j)]\,.$$

If $i=j$, we get no contribution to the sum. If $i\neq j$, the terms involving i and j give us the expression

$$\frac{m_im_i}{l_i^3\,l_j^3}[(y-y_i)^2(x-x_j)^2-(x-x_i)(x-x_j)(y-y_i)(y-y_j)$$
$$+(y-y_j)^2(x-x_i)^2-(x-x_i)(x-x_j)(y-y_i)(y-y_j)]\,.$$

But this is exactly

$$\frac{m_im_j}{l_i^3\,l_j^3}\{(y-y_i)(x-x_j)-(y-y_j)(x-x_i)\}^2\,.$$

Lemma 2. *If not all of the points P_i are on the same straight line, then $F_{xx}F_{yy}-F_{xy}{}^2>0$ for all points $P\neq P_i$, $i=1,\ldots N$.*

This follows directly from Lemma 1.

For the next lemma, we consider a straight line S passing through the point $Q(x_0,y_0)$. Thus, $S(t)=[x(t),y(t)]$ is the parametric representation of the line where $x(t)=x_0+at$ and $y(t)=y_0+bt$. Let h be the function defined by the formula $h(t)=FoS(t)=F[x(t),y(t)]$.

Lemma 3. *F has no maximum value at any point $Q\neq P_i$, $i=1,\ldots N$.*

Proof: Suppose F has a maximum at $Q(x_0,y_0)$. Then let $h(t)=F(x(t),y(t))$ where x and y are defined as above. Then h has a maximum at 0. But

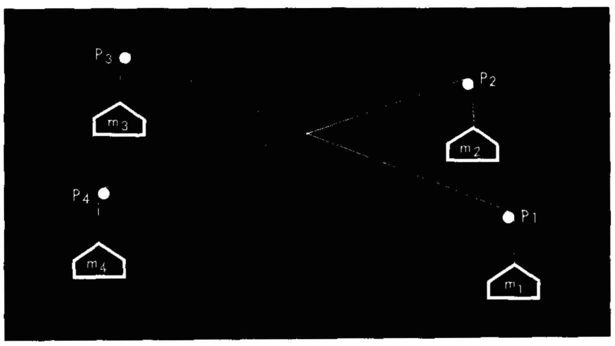
$$(1)\quad h'=F_x\left(\frac{dx}{dt}\right)+F_y\left(\frac{dy}{dt}\right)$$

$$(2)\quad h''=F_{xx}\left(\frac{dx}{dt}\right)^2+2F_{xy}\left(\frac{dx}{dt}\right)\left(\frac{dy}{dt}\right)+F_{yy}\left(\frac{dy}{dt}\right)^2.$$

These derivatives exist at all points $Q\neq P_i$.

Then clearly, $h''(t)>0$ for all t where h'' exists. This follows from Lemma 2, because the discriminant of the quadratic form (2) is negative. Thus $h''(0)>0$ for all t where h'' exists, and $h'(0)=0$. Also $h(0)=F(x_0,y_0)$. Thus, h has a local minimum at 0, contradicting the hypothesis that h has a maximum at 0.

Figure 1



Lemma 4. *If S is a straight line through P_i, then $h=FoS$ has no maximum at P_i.*

Proof: Let $\left.\begin{array}{l}x(t)=x_i+at\\ y(t)=y_i+bt\end{array}\right\}$ be parametric equations for the line S.

Then $h(t)=\sum_{j\neq i}m_jl_j(t)+m_i\sqrt{(x-x_i)^2+(y-y_i)^2}$

$=h_0(t)+m_i|t|\sqrt{a^2+b^2}$,

where $h_0(t)=\sum_{j\neq i}m_jl_j(t)$.

h_0 can have no maximum at $t=0$ by Lemma 3. $m_i|t|\sqrt{a^2+b^2}$ has a minimum at $t=0$.

Theorem 1. *F can have at most one minimum on the plane.*

Proof: Suppose F has two minima, say at P and Q, respectively. Then let S be a straight line passing through P and Q, such that $S(0)=P$, $S(1)=Q$. Then consider $h\equiv FoS$: $[0,1]\to R$. By the Weierstrass theorem, h has a maximum at some point t_0 in $(0,1)$ because $h(0)$ and $h(1)$ are minima. This contradicts either Lemma 3 or 4.

Theorem 2. *The function F has at least one minimum.*

Let l_{ij} be the distance between P_i and P_j and choose K so that $K>\max_{i,j} l_{ij}$. For any point Q, we let l_{iQ} be the distance from P_i to Q and $d(P,Q)$ be the distance between P and Q. Then for any point Q such that $l_{1Q}\geq 2K$, we have $l_{iQ}>K$ for all i. From the triangle inequality, we have $l_{iQ}\geq l_{iQ}-l_{1i}\geq 2K-l_{1i}>K$.

Thus $F(Q)=\Sigma m_il_{iQ}>\Sigma m_iK>\Sigma m_il_{i1}=F(P_l)$. (1)

Let C be the set of points P such that $d(P,P_1)\leq 2K$. C is a bounded closed set, and since F is continuous, by the Weierstrass theorem, it has a maximum on the set C.

Let $\overline{P}$ in C be a point where $F(\overline{P})$ is minimum.[1] Then $F(\overline{P})$ is the minimum for the function F. For if Q is any point in the plane, $Q=C$ or $Q\neq C$. If $Q=C$, then $F(\overline{P})\leq F(Q)$ because $F(\overline{P})$ is the minimum over C. If $Q\neq C$, then $d(Q,P,)>2K$, hence $F(\overline{P})\leq F(\overline{P_1})<F(Q)$.

Combining Theorems 1 and 2, we have proved that the function F has exactly one minimum.

Analog method

Another approach to the problem of finding the minimum of F is the use of the following analog[3] device (Fig. 1). It is a generalization of the method described by Polya[2] for the treatment of a similar problem.

On a map mounted on a wooden board, a pulley is attached at each station P_i, and n strings are tied together in a knot at a common point. At the other end, a weight proportional to m_i is attached and the string is suspended over the pulley at the point P_i. After all of the weights are attached, the board is held in a vertical plane. The desired point C is the location of the knot when the system is allowed to hang freely. The pulleys are used to reduce the friction and the board is jiggled so that sticking does not occur.

Note that C is the point where the potential energy of the system is a minimum. Let L_i be the length of the string which is hung over the pulley at point P_i. Let s_i be the length of string from the pulley to the weight, h_i the distance of the weight to a fixed horizontal reference plane, H_i the distance from the pulley to the fixed horizontal plane, and l_i the distance from the pulley to the knot. Then the expressions Σm_iH_i and Σm_iL_i are both constants. Also we have

(1) $\Sigma m_iH_i = (\Sigma m_ih_i) + \Sigma m_is_i)$

(2) $\Sigma m_iL_i = \Sigma m_is_i + \Sigma m_il_i$.

Thus, subtracting (1) from (2) we obtain

$$\Sigma m_il_i - \Sigma m_ih_i = (\Sigma m_iL_i - \Sigma m_iH_i)$$

$$\text{or } \Sigma m_il_i = \Sigma m_ih_i + (\Sigma m_iL_i - \Sigma m_iH_i)\,.$$

Since the last two terms are constants, we see that the function Σm_il_i attains a minimum whenever the function Σm_ih_i attains a minimum, but this last function is just the potential energy of the system of weights. In this method it turns out that the center is that point where the forces corresponding to the weights are in equilibrium.

It is interesting to note that the location of the center is not the center of mass, as one might think upon first meeting this problem. The case of three noncollinear stations with equal weights illustrates this. Here the center is that point for which the line segments to the three vertices form angles of 120° with each other, if this point is inside the triangle (Fig. 2). If this point is outside the triangle, then the center is at the vertex with the obtuse angle (Fig. 3).

Conclusion

The principal result of this paper is the proof that there is one and only one point for the center which gives a minimal network when all of the stations are to be connected directly to the center. Thus one can use any search procedure to find the coordinates of the center.

Although this treatment is only a solution of a special case of the problem of obtaining the optimal network, we have found by calculation on a real network that substantial savings of line costs can be realized. The optimal network problem is extremely complicated. However, efforts are under way to use and extend the methods described here to devise means for its treatment. The simple nature of the analog procedure suggests that clever use of this technique may be extremely useful in complex problems of this type.

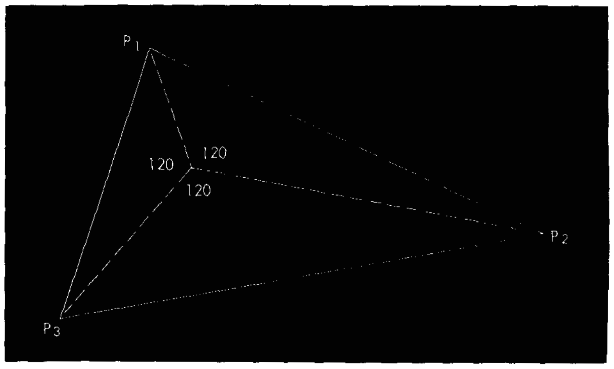


Figure 2

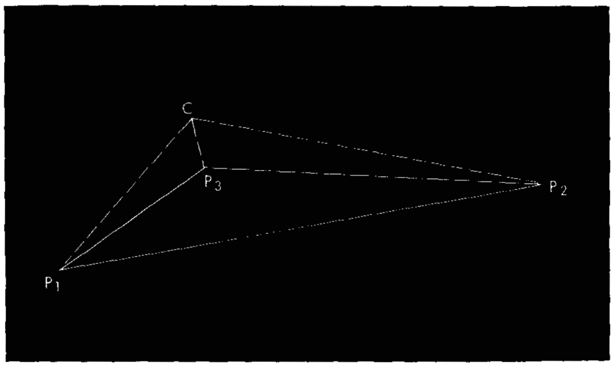


Figure 3

Acknowledgment

I would like to thank Geza Schay and Jay M. Berger for many interesting conversations and discussions relating to the subject of this letter.

References and footnotes

1. $\bar{P}$ is not on the boundary of C, because if it were, then $d(\bar{P}, P_1) = 2K$ and by (1), $F(\bar{P}) > F(P_1)$, contradicting the minimal nature of $\bar{P}$.
2. G. Polya, *Mathematics and Plausible Reasoning,* Vol. 1, Princeton University Press, 1954, pages 147-8.
3. W. Miehle, "Link-Length Minimization in Networks," *Operations Research,* **6,** No. 2, 232-243 (1958). Other interesting analog methods are discussed in Miehle's paper.

Received July 25, 1961