

# Location Optimization Problem with Multi Refineries and Multi Stations

Na Wang<sup>1</sup>, Xinshe Qi<sup>1</sup>, Jianhang Zhang<sup>2</sup>, Qingzheng Xu<sup>3</sup>

<sup>1</sup> Department of Basic Courses, Xi'an Communications Institute, Xi'an, 710106, China

<sup>2</sup> Department of Information Security, Xi'an Communications Institute, Xi'an, 710106, China

<sup>3</sup> Department of Information Service, Xi'an Communications Institute, Xi'an, 710106, China

**Keywords:** location optimization problem, differential evolution algorithm, clustering algorithm, construction cost.

**Abstract.** In this paper, the contest problem C of 2010 Contemporary Undergraduate Mathematical Contest in Modeling is extended to location optimization problem with multi refineries and multi stations, and the solution procedure is divided into two main phases. In the first stage, the objective function of construction cost is established under the distribution of multi refineries and the unit cost of pipeline, and then the extremal solution is obtained by differential evolution in view of the connection status of some refineries. In the second stage, the exhaust algorithm, which has high complexity, is discussed to provide the exact solution. After that an improved clustering algorithm is proposed to directly define the category of multi refineries. The experimental results show that, the proposed approach can greatly improve computation efficiency without severely sacrificing the accuracy of the results, which especially meets the objective requirement of engineering practice.

## 1. Introduction

In a word, the purpose of location optimization problem is to decide a suitable location for infrastructure and then obtain some optimal solutions in some special cases. The infrastructure here includes manufactory, station, first-aid centre, firehouse, garbage disposal center, logistics center, and ammunition depot. As a classical problem in operational research, the location optimization problem has been widely and successfully used in daily production and life, logistics and transportation, and even military operation. A good result of location optimization can not only reduce service cost and inventory but also improve efficiency and cost-effectiveness. In capital market, the total profit and market share of the company are expected to rise at a considerable rate. So, the research of location optimization problem has a significant economic and social role in present-day.

The Contemporary Undergraduate Mathematical Contest in Modeling (CUMCM) is an annual contest for undergraduates all over the world, organized by China Society for Industrial and Applied Mathematics. As a special type of location optimization problem, the contest problem C of CUMCM-2010 is how to plan out the oil pipeline for petroleum enterprises [1]. Generally speaking, the layout of oil pipeline is determined by natural factors, economic factors and social factors jointly. To maximize the economic efficiency of enterprises, we should consider carefully many influencing factors, such as the distance between the petroleum refinery and the station, the unit expense and the quality of pipeline, the laying cost of different geographical areas, the demolition compensation of different areas, and the topology of the existing pipeline network.

With the help of theoretical analysis and simulation tools, several people have done much research from different angles and obtained some fruitful outcomes based on different mathematical models in recent years. Seen from the research objects, the original problem was successfully solved by most researchers in order to obtain more precise solution with more laconic method. What's different about that is the original problem under different geographical areas was expanded for more general situation by Song et al. in 2011 [2]. Based on the assumption that there are some given relations between the demolition compensation of the city proper and the distance to the suburbs, a more general mathematical model was then set up to minimize the layout cost of the pipeline.

To divide from the research approaches, there are four classes to solve this contest problem until now. (1) Mathematical software method [3-6]. Some researchers constructed a mathematical model, and then found one or a group of numerical solutions by MATLAB, Mathematica, LINGO or other specialized software. Although they can calculate the final output values quickly by this method, they can not connect the results with their value for practical problems, as just few data is always not meaningful for decision-makers in practice. (2) Geometry method [7]. The core part of original location optimization problem can, indeed, be translated into a weighted version of Fermat's problem of points. An intuitive and effective way to solve the classical Fermat's problem is by using elementary geometry. Unfortunately, often relied the given problem, this geometry method is not suitable for most engineering applications. (3) Analysis method [8-12]. A binary function is structured firstly, in which the variables are the geometry coordinates of the joint point between the shared pipeline and the non-shared pipeline. Then a set of fixed-point equations is got by means of the extremum theory of function of several variables. Solving this set of equations, the key and difficult point of this method, needs subtle and imperceptible skills. According to latest studies as far as we know, several methods were proposed to solve the set of equations, such as replacing method by trigonometric function, method of undetermined coefficients, direct deformation method, and mathematical software. (4) Combination method of geometry and analysis [2, 13-16]. The least cost function of one variable is build with the help of plane geometry firstly. And then the objective function is easily solved based on the properties of the function, such as monotonicity and convexity. This method is easy to learn for its concise procedure, yet it is very hard to expand to other related optimization problems.

The location optimization problem for two-dimensional system is discussed for more extensive background in this paper. For the contest problem C of CUMCM-2010, the geometric coordinates of two oil refineries are given as a known condition and a new train station should be fixed in order to minimize the construction cost. Naturally, we want to know that how to build  $n$  new stations for the least cost of the pipeline construction, when  $m$  oil refineries are known in advance. For simplicity, these optimization problems are denoted by  $P(m, n)$ . To the best of our knowledge, the extended optimization problems have not yet been studied and solved successfully by other researches. It is obvious that the contest problem C of CUMCM-2010 is only a special case of  $P(m, n)$ , in which  $m = 2$  and  $n = 1$ .

According to common sense, when  $m \leq n$ , that is the number of oil refineries is less or equal to the number of stations,  $m$  stations are coincident with the feet of the perpendiculars from  $m$  oil refineries to the railway line. Thus it can be addressed easily as a common problem. Hence in this paper we will make research mainly on the optimization problems in which  $m > n$ .

For the extended optimization problem, it is difficult to determine the objective function and its properties are more diverse than previously described. As a result, applying classical theory deduction, the difficulty of solving such problems has increased dramatically and it is totally infeasible under some extreme cases. So we shall follow the operation strategy of "easy first and difficult later" in this paper. That is to say, the problem of  $P(m, 1)$  is solved firstly, and then, based on previous approach, we will, step by step, tackle the problem of  $P(m, n)$ . During the computational produce, with the help of a well-known intelligent algorithm, some approximate optimal solutions can be obtained under limited computational time and memory resource in order to meet demand of engineering applications.

## 2. Approach of Problem $P(M, 1)$

### 2.1 Theoretical Discussion

As illustrated in Fig. 1(a), the number of oil refineries is obviously expanded from 2 of the original contest problem C of CUMCM-2010 to  $m$  of the problem  $P(m, 1)$ . It is supposed that, the coordinates of oil refineries are  $(p_i, q_i)$  ( $i = 1, 2, \dots, m$ ), the coordinates of new station E is  $(x, 0)$ , the coordinates of start point M of shared pipeline is  $(x, y)$ , and the unit pipeline expense of non-

shared pipeline and shared pipeline are  $c_i$  ( $i = 1, 2, \dots, m$ ) and  $c_0$ , respectively. Then, the objective function, the cost of the pipeline construction, can be abstracted as the following

$$F(x, y) = \sum_{i=1}^m c_i \sqrt{(p_i - x)^2 + (q_i - y)^2} + c_0 \cdot y \quad (1)$$

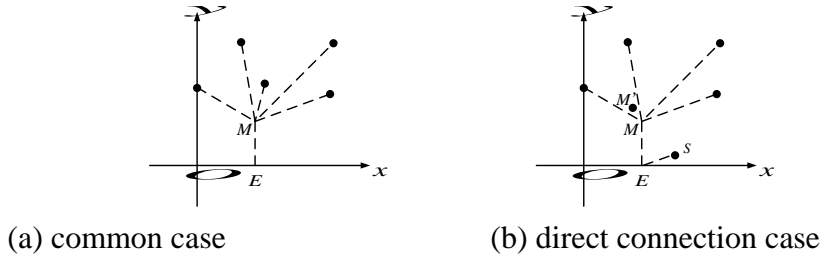


Fig.1: Problem P(m, 1)

Compared with the original contest problem, the expression of objective function has not changed much. Then the partial derivatives to  $x$  and  $y$  can be calculated respectively. According to calculus theory, the possible optimal value must be obtained if and only if the partial derivatives are equal to zero at the same time. Along with the number of oil refineries  $m$ , the complexity of objective function is increasing inevitably. Thus it is easy to fall into the curse of dimensionality and difficult to solve the set of binary equations. Hence we may conclude that the classical analysis basically loses effectiveness, and in my opinion, it has more value in theory than in its applications.

To solve the complicated multi-function optimization problem, some novel intelligent methods are proposed, such as genetic algorithm (GA), particle swarm optimization (PSO), artificial neural networks (ANN), and differential evolution (DE). Compared with classical methods, these intelligent methods have little requirements on objective function, and they are valid especially to high dimension and complex function optimization problem. DE is proposed by Storn and Price, which is an effective, robust, reliable, and simple global optimization algorithm [17]. According to frequently reported comprehensive studies [18-21], DE has been preferred to many other optimization techniques such as GA, PSO, evolutionary programming (EP), and simulated annealing (SA) in terms of accuracy, convergence speed, computation complexity, and robustness over both well-known benchmark functions and hard real-world problems. So in this paper, DE algorithm is utilized to solve the location optimization problem modeled previously. Here the flow of DE algorithm is not stated clearly. For further details please carefully read the reference cited in this subsection.

In further experiment, we find an important question to be answered is whether all pipelines from all oil refineries have to converge at point  $M$  and then connect with point  $E$  together. The answer to this question is NOT in my own mind. In other words, a part of oil refineries may intersect with the railway line directly. As seen from Fig. 1(b), the oil refinery  $S$  is connected with point  $E$  on the railway line, instead of point  $M$  firstly. It is intuitive from Fig. 1(b), that the construction cost may be less than previous case shown in Fig. 1(a). Different with Fig. 1(a), this special example of optimization problem in this paper is called problem P(m, 1) in direct connection case.

Accordingly the next question is that which one or more of oil refineries are connected with station (foot of the perpendicular on the railway line) directly. To answer this question, we will carefully discuss the variation tendency of the cost of the pipeline construction under two different situations. It is obvious that, for most oil refineries, the construction cost will not change any more because their connection mode is fixed by comparing Figs. 1(a) and 1(b). As a result, the construction cost of point  $S$  plays a decisive role in calculating the total construction cost, and then it alone is considered in next analysis. For the common case as seen from Fig. 1(a), the construction

cost of point  $S$  is equal to  $c_s \sqrt{(p_s - x)^2 + (q_s - y)^2}$ , and it is equal to  $c_s \sqrt{(p_s - x)^2 + q_s^2}$  for the direct connection case from Fig. 1(b). So it is concluded that, the construction cost in direct

connection case is less than that in common case if  $q_s < \frac{y}{2}$ , and the result is just the reverse if  $q_s > \frac{y}{2}$ . In addition, it is worth noting that, the vertical coordinate  $y$  of point  $M$  is calculated based on the coordinates of all oil refineries. For the common case, the point  $S$  is included in these oil refineries. What's different about this point is that, it is not included for the direct connection case. For the sake of simplicity, the minor differences between them would be laid aside, and all vertical coordinates are calculated in the situation which includes the point  $S$  in this paper.

## 2.2 Numerical Example

Based on the algorithm analysis, we developed a program for this optimization problem. The algorithm flow chart is as follows in Fig. 2.

The validity of this algorithm is then tested on problem  $P(5, 1)$ . That is to say, there are five oil refineries to be built on the same side of a railway. We make a hypothesis that, the coordinates of these oil refineries are determined as  $(20, 30)$ ,  $(70, 40)$ ,  $(5, 50)$ ,  $(30, 45)$  and  $(90, 5)$ , the laying cost of the shared pipeline is 7.2, and the cost of the non-shared pipeline from each oil refinery are 5.6, 6, 6.2, 5.4 and 6.9, respectively. For DE algorithm, differential amplification factor  $F$  is 0.5, crossover probability constant  $Cr$  is 0.9, population size  $PopSize$  is 50, and maximum NFC  $MaxFunEva$  is 2000 in our experiments.

With the help of MATLAB R14, a mathematical software developed by Mathworks Inc., we obtain all the data we needed. The main experimental results are the coordinates of joint point  $M$  and station  $E$  in railway are  $(34.1828, 34.4189)$  and  $(34.1828, 0)$ , respectively, and the total cost  $F(x, y)$  is equal to 1201.7. Seen from the connection relation, the pipelines from the first four oil refineries converge at joint point  $M$  firstly and then intersect with the railway line at station  $E$ . The fifth oil refinery, a special point, is yet connected directly with station  $E$ .

In fact, if all five oil refineries converge at joint point firstly and then intersect with the railway line, the data we hold are the coordinates of joint point  $M$  and station  $E$  are  $(33.7847, 30.4639)$  and  $(33.7847, 0)$ , respectively, and the total cost  $F(x, y)$  is equal to 1243.9. It is obvious that the laying cost in this situation is higher than that situation described above, which can show, from one aspect, that the algorithm proposed in this paper is correct.

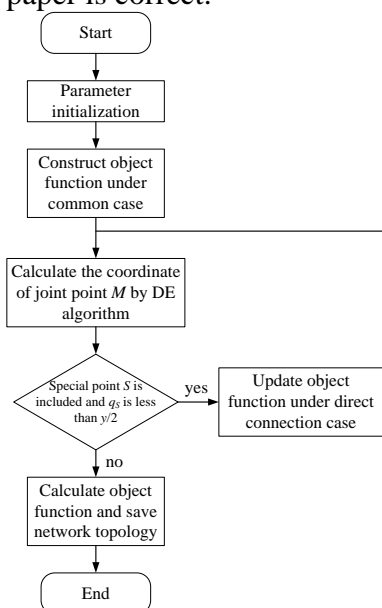


Fig.2: The algorithm flow chart

### 3. Approach of Problem P(M, N)

#### 3.1 Theoretical Discussion

Next, we will discuss the optimization problem  $P(m, 2)$  on the basis of analysis above. However, the following new problem exposes clearly in this discussion. Which one or more of oil refineries are connected with station 1 in order to transport oil? The same question is also for the stations 2, 3, and so on. It is supposed that, for each station used to transport oil, there are more than one oil refineries connected with it (the same below). Then there are  $m-1$  division possibilities for problem  $P(m, 2)$ , and even  $C_{m-1}^{n-1}$  division possibilities for problem  $P(m, n)$ .

For every division possibility, problem  $P(m, n)$  can be converted to  $n$  problems:  $P(m_1, 1)$ ,  $P(m_2, 1)$ , ..., and  $P(m_n, 1)$ , where  $m_1 + m_2 + \dots + m_n = m$ . Naturally, these simple optimization problems can be solved by the approach described in Section 2.1. Finally, we can select an optimal answer from  $C_{m-1}^{n-1}$  candidate solutions, and then, as illustrated in Fig. 3, the difficult problem  $P(m, n)$  is successfully solved in this way.

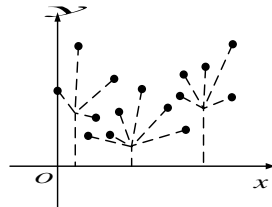


Fig.3: Problem  $P(m, n)$

#### 3.2 Engineering Applications

Firstly, when the exhaust algorithm described in Section 3.1 is applied to solve problem  $P(m, n)$ , the intelligent algorithm are executed  $n \times C_{m-1}^{n-1}$  times in this course and each process of intelligent algorithm may consume lots of memory space and time. It is undoubted that the amount of work will be very huge with the increasing of  $m$  and  $n$ . From another viewpoint, every population-based intelligent algorithm can not, in theory, guarantee the optimal solution for all time. So even an optimal solution is obtained by the exhaust algorithm, it can not guarantee a globally optimal solution in this approach.

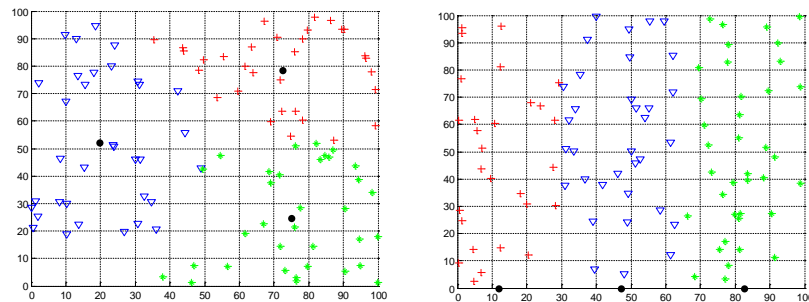
Based on the two considerations above, we have to take the second best option. That is to say, based on the idea of clustering algorithm,  $m$  oil refineries are allocated reasonably to  $n$  stations at one time. Now the intelligent algorithm are only executed  $n$  times in this way. As a result, we can greatly improve computation efficiency without severely sacrificing the accuracy of the results.

For its terseness and efficiency, K-means clustering is recognized as one of the simplest unsupervised learning algorithms that solve the well known clustering problem and has been successfully adapted to various problem domains, ranging from market segmentation, computer vision, geostatistics, and astronomy to agriculture [22, 23]. For further details of the classical K-means clustering, please read the reference cited in this subsection.

Generally speaking, when this clustering algorithm is employed to define centroids for all clusters, Euclidean distance in two dimensions is often a chosen distance measure and then  $k$  centroids are distributed inevitably in the full plane. For example, the clustering result by the K-means algorithm is shown in the following Fig. 4(a), where 100 points are distributed in two-dimensional plane randomly.

But in this paper, the station must be located at the railway line ( $x$  axis in Fig. 4), not other locations in the full plane. Therefore, in K-means algorithm, Euclidean distance in one dimension may be a suitable distance measure between data points. In one dimension, the distance between two points in  $x$  axis is the absolute value of their numerical difference. Based on the analysis and discussion, an improved K-means clustering algorithm is proposed in this paper. The clustering

result by the improved K-means algorithm is illustrated in Fig. 4(b), where 100 points are also distributed in two-dimensional plane randomly. This time, the coordinates of three centroids are (11.7627, 0), (47.1705, 0), and (82.5925, 0). Meanwhile, 100 points are allocated evenly to different clusters.



(a) Euclidean distance in two dimensions (b) Euclidean distance in one dimension

Fig.4: K-means clustering algorithm

After carefully observation and analysis, we can find that the points, which horizontal ordinates are below than 29.4666, are bracketed into the first cluster, the points, which horizontal ordinates are above than 29.4666 and below than 64.8815, are bracketed into the second cluster, and the points, which horizontal ordinates are above than 64.8815, are bracketed into the third cluster. Here, 29.4666 is equal to  $(11.7627+47.1705)/2$ , and 64.8815 is equal to  $(47.1705+82.5925)/2$ . From a geometric perspective, the relation of each random point and its corresponding centroid is uniform. The points which fall in between two adjacent centroids are divided by the geometric centre. Particularly, the points in left side are belong to left centroid, and the points in right side are belong to right centroid. Obviously, this conclusion and experiment result accord with common sense.

#### 4. Conclusion

Striving for more universality, the contest problem C of CUMCM-2010 was extended to location optimization problem with multi refineries and multi stations. This extended optimization problem was difficult to solve using previous methods, especially for large scale applications. In this paper the procedure of solving this difficult optimization problem was divided into two main phases. At the first stage, problem  $P(m, 1)$  was solved firstly by differential evolution algorithm, specially considering all connection relations of oil refineries and stations. At the second stage, for solving problem  $P(m, n)$ , an improved clustering algorithm instead with distance measure in one dimension, was proposed to allocate the oil refineries reasonably. The method could greatly improve computation efficiency without severely sacrificing the accuracy of the results. And then the contradiction between workload and accuracy was tracked skillfully, which meets the objective requirement of engineering practice.

In the near future, the classical differential evolution algorithm and K-means clustering algorithm can be modified to improving the efficiency. Furthermore, the special point (point S in this paper) can included or not, when the geometric center point M is calculated for the common case of problem  $P(m, 1)$ . However the geometric center point found may vary slightly or seriously in different situations. So our top task in the future is how to consider the small difference between them and then design new algorithm more reasonably.

#### Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 61305083).

## References

- [1] National Organizing Committee of the Contemporary Undergraduate Mathematical Contest in Modeling, CUMCM-2010 contest problems, available at <http://en.mcm.edu.cn/>.
- [2] Song B., Cao J.D. & Chen E.S., An optimizing model for oil pipe arrangement with changeable removal compensation. *Mathematics in Practice and Theory*, 41(20), pp.248-252, 2011. (in Chinese)
- [3] Jia C.J., Pipeline layout scheme of two oil refineries based on nonlinear programming — Case of mathematical contest in modeling solutions for 2010 higher education community cup. *Journal of Yueyang Vocational Technical College*, 25(5), pp.61-65, 2010. (in Chinese)
- [4] Liu W.B., Fan K.T., Song C. & Wang Y., Make arrangements for petroleum pipe line. *Mathematical Theory and Applications*, 31(4), pp.72-78, 2011. (in Chinese)
- [5] Zhao S.Z. & Zhang T., The mathematic pattern and solution on arrangement of oil transportation pipelines. *Journal of Gansu Lianhe University (Natural Sciences)*, 25(6), pp.20-23, 2011. (in Chinese)
- [6] Tang Z.H., The mathematical mode of oil pipeline design and construction cost. *Journal of Gannan Normal University*, 31 (6), pp.28-32, 2010. (in Chinese)
- [7] Mao J.S., The analysis of examination questions of mathematical contest in modeling — Design scheme of oil delivery pipes. *Journal of Luzhou Vocational Technical College*, 26 (1), pp.61-64, 2011. (in Chinese)
- [8] Gao C.C., Qi L.H., Chen L.J. & Li Y.N., The optimal schemes of pipe laying. *Pure and Applied Mathematics*, 27(3), pp.411-414, 2011. (in Chinese)
- [9] Wu L., Optimization model for the layout of pipeline. *Journal of Changshu Institute Technology (Natural Sciences)*, 25(10), pp.58-61, 2011. (in Chinese)
- [10] Qi H.P. & Zhao X.M., The arrangement of the pipelines project. *Journal of Gansu Lianhe University (Natural Sciences)*, 26(1), pp.27-30, 2012. (in Chinese)
- [11] Wang N., Xu Q.Z. & Zhang Y.F., Mathematical modelling and solution of pipeline laying problem. *International Journal of Simulation and Process Modelling*, 7(1/2), pp.16-22, 2012.
- [12] Xin K.D. & Wang H., Oil pipelines decorating model and application. *Journal of Yili Normal University (Natural Science Edition)*, 7(1), pp.17-23, 2013. (in Chinese)
- [13] Sun Y.P. & Cao X.L., The mathematical model of the pipeline layout. *Journal of Changzhou Institute of Technology*, 24(2), pp.32-37, 2011. (in Chinese)
- [14] Lu W.F. & Chen W.Z., Mathematical model of optimal arrangement of petroleum pipeline. *Journal of Suzhou Vocational University*, 22(3), pp.60-65, 2011. (in Chinese)
- [15] Pan L.J., Optimization model of pipeline arranges. *Journal of Weinan Teachers University*, 26(6), pp.18-21, 2011. (in Chinese)
- [16] Dong X.Y., Optimization model of pipeline arranges. *Science Technology and Engineering*, 24(11), pp.5890-5892, 2011. (in Chinese)
- [17] Storn R. & Price K., Differential evolution – A simple and efficient adaptive scheme for global optimization over continuous spaces. Berkeley, CA, Tech.Rep.TR-95-012, 1995.
- [18] Andre J., Siarry P. & Dognon T., An improvement of the standard genetic algorithm fighting premature convergence in continuous optimization. *Advance in Engineering Software*, 32(1), pp.49-60, 2001.
- [19] Hrstka O. & Kucerova A., Improvements of real coded genetic algorithm based on differential operators preventing premature convergence. *Advance in Engineering Software*, 35(3/4), pp.237-246, 2004.
- [20] Vesterstrom J. & Thomsen R., A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. in *Proceeding of Congress on Evolutionary Computation*. Portland, Portland USA, 1980-1987, 1987.
- [21] Price K., Storn R.M. & Lampinen J.A., *Differential Evolution: A Practical Approach to Global Optimization (Natural Computing Series)*. Springer-Verlag, New York, USA, 2005.

- [22] MacQueen J.B., Some methods for classification and analysis of multivariate observations. in Proceeding of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, USA, 281-297, 1967.
- [23] Anil K.J., Data clustering: 50 years beyond K-Means. Pattern Recognition Letters, 31(8), pp.651-666, 2010.