Mathematical Problems of Computer Science 23, 2004, 5–10.

Data Flow Analysis By Linear Programming Model

Armen H. Alaverdyan, Vilik M. Karakhanyan and Vardan R. Tonoyan

Institue for Informatics and Automation Problems of NAS of RA
e-mail armen_am@yahoo.com

Abstract

The paper presents the general discussion of the data flow algorithms structuring problem on one hand, and, as an example of that particular problem class, the analysis of the linear programming problem when the objective function coefficients vary depending on the data flow. The problem is in reconstruction of current result with such an approach, which is the most plain from the full solution of the problem by the dynamic data set flow.

1 Introduction

Many modern applications of information technologies, such as network monitoring, optimal management of telecommunication systems, network search, the exploitation of network measuring instruments and their data bases, etc., deal with continuous data flows and unusual, non-finite and nonstoraged data set. In this case, the requests (the requirements of data analyses) in their turn are longterm and continuous processes in contrast to usual one-time questions. The traditional data bases and the traditional data processing algorithms are poorly adjusted to the processing of difficult and continuous requests in the data flows. This generates a necessity for new investigations for continuous, multipart, depending on time and subjected to indefinite behavior of data flows processing [1]. Concerning the mentioned class, some systems and algorithms are processed for different needs: real-time working systems, controlling systems, modelling processes, etc., but they are episodic for the formulated general problem.

The incremental update algorithmic model of data analysis by changed problem conditions prefers episodic change of problem solution to its full analysis [2]. For example, a particular change of time-table problem conditions must be brought to time-table episodic reconstruction. It is obvious that it is possible to construct theoretical problem where any particular change brings to the full reconstruction of the problem. It is also clear that there are numerous problems which are not critical to the local transformations. It is possible to solve any problem by mentioned algorithms, moreover, in the specific conditions it is the only possible variant for solving a problem, including the data flows analysis.

The present statement will discuss an important applied model in the flows environment. We are going to consider a linear programming mathematical problem the parameters of which are formed by data flows. At the moment it is assumed that the optimal plan is found and the coordinates of objective function vary. In this case, there is an emerging question of the mechanisms by which it is possible to follow the coefficients variation by creating the next configuration of the change optimal plan. It is clear that this relates to simple change of the current plan, but not the full analysis of the problem.

2 Linear Programming Problem In The Flows Environment

Particular examples of linear programming general problem are given through the definition of coefficient values: a_{ij}, c_j, b_i $(i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n)$. Let's examine a linear programming problem of canonical view:

$$\min\ \mathbf{c}'\mathbf{x},$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq 0,$$

where $\mathbf{c} \in R^n$, $\mathbf{b} \in R^m$, $\mathbf{A}$ is the $m \times n$ full rank matrix, $m < n$. Let's imagine the peculiar situation arising from applications in which the mentioned coefficients are changing. Such problems appear, for example, for data flows.

$B(t, n)$ data flow is finite but a very large-sized sequence of $b_1, b_2, ..., b_n$, where $b_t, t = \overline{1, n}$ are certain structures. The data flows processing algorithms can use incomparably smaller storage memory than the input data. The limitation window is given in certain cases for processing separate data episods. The time-dependent values of the applied model parameters are obtained as result of analysis. Unlike other natural and similar definitions, the variation of parameters is unpredictable here, as it does not have probabilistic distribution and is not described by one or another property. Instead, it is considered that the variation takes place very slowly, because of the accumulation. In turn, the applied problem demands to have ready answers to the certain questions.

There are two strategies: to solve a problem for every moment by all present data which is practically impossible because of the data sizes, and to structure growing generation systems when the new data analyses are more or relatively easy summarized with the results of the previous data analyses.

We are going to consider linear programming in the mentioned conditions. Any variation of the coefficients is not considered. Instead, insignificant and slow variation is considered so that the variation is fully monitored and it changes the solution point slowly. Of course, it is possible to formalize this fully. At the same time, it is possible to consider partial behavior of parameters variation by providing clarity of analysis and getting optimum change mode description in simple terms. The clarification is also expedient in respect to the method application. Let's suppose that in $Z = \sum_{j=1}^{n} c_j x_j$ linear objective function of the linear programming problem c_j coefficients are variated by $B(t, n)$ flow. Assume that t_0 is the fully analized moment, i.e. we know about the existence of optimization at that moment and the vertex, if the latter exists. This vertex is stable for certain variation of c_j coefficients [3]. The stable set is described by simple inequalities and it is clear that there is a necessity to consider its borders. The theoretical analysis of it is presented in [4] which proves that the feasible polyhedron vertices set is divided into equivalent vertices groups and that the passage from one optimization vertex to another takes place through those groups. Some of the mentioned groups are neighboring with a separately taken vertex, and the total sum of those groups vertices can be large. Theoretically, in terms of flows, having an optimization vertex, it is necessary to prepare neighboring equivalent vertices by calculating current c_j coefficients. The weakness of this approach direct application is the drastic increase in the number of calculations for decreasing which approximations and heuristic solutions should be applied. The below considered natural approach gives primary significance to the vertices which are adjacent to the variation process.

At t_0 moment let's denote the optimal vertex by $\bar{x}^{t_0}$ and $\bar{c}^{t_0}$ is the corresponding vector of coefficients. Let's review $\bar{c}^{t_0}$ and $\bar{c}^{t}$. It is clear that variation in the approach from $\bar{c}^{t_0}$ to $\bar{c}^{t}$ is arbitrary and it is controlled by flows. It is important that during the variations vector of coefficients $\bar{c}^{t}$ do not go out $\bar{x}^{t_0}$ vertex oprimization area. It is possible to apply a set of simple mechanisms for further behavior modelling of $\bar{c}^{t}$. For example, one may consider a standard deviation of $\bar{c}^{t_0}$ from$\bar{c}^{t}$ direction. The most simple model is the consideration of that direction. As an extrapolation, it leads to the intersection of $\bar{x}^{t_0}$ vertex and the hull of stability at the most probable point. In case of sufficient resources, it is also possible to consider some vicinity of that point, but it is important that in contrast to the mentioned theoretical model these application approaches give an opportunity to work with the limited quantity of the possible vertices. Depending on the cosidered problem parameters, the algorithmic system is able to choose a scheme corresponding to extrapolation, which deals with different numbers of neighboring vertices. The basis of the choice by $\bar{c}^{t}$ flow average measures and dispersion is in the process of calculation, which is a simple flow problem (it is showed in the [1]). Assuming that this question is clarified, let's consider the problem behavior in the case of rectilinear variation.

3 Rectrilinear Extrapolation

In the above mentioned case, variation of the objective function coefficients is more expedient to describe with $c_j(\lambda) = c_j^{t_0} + \lambda(c_j^{t} - c_j^{t_0})$ expression, where λ varies in the certain limits, $[0,1]$ interval is internal and characterizes the variation from $\bar{c}^{t_0}$ to $\bar{c}^{t}$, and $\lambda > 1$ values extrapolate the further values of coefficients vector in the line of variation. Let's denote $c_j^{\Delta} = c_j^{t} - c_j^{t_0}$.

So, a linear fuction

$$Z = \sum_{j=1}^{n} (c_j^{t_0} + \lambda c_j^{\Delta}) x_j (1)$$

and the system of linear limitations

$$\begin{array}{ll} \sum_{i=1}^{n} a_{ij} x_j = b_i, & i = 1,2,\ldots,m, \\ x_j \geq 0, & j = 1,2,\ldots n, \end{array} \quad (2)$$

are given.

It is nessecary to convoy λ change and find out from $1 < \lambda$ interval the minimal value of λ at which the variation of the optimization vertex takes place for the first time. Assume that the vector $\bar{x}^{t} = (x_1^{t}, x_2^{t}, \ldots, x_n^{t})$, which satisfies system (2), introduces corresponding new optimization basis.

According to the hypothesis, in case of $\lambda = 0$ initial value, we have an optimal solution. Assume that some basis (consisting of the m vectors of $\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_n$ system) corresponds to it. According to the simplex algorithm optimization condition, its all estimations in this case must correspond to the following condition: $z_j - c_j^{t_0} \leq 0$. As $c_j(\lambda) = c_j^{t_0} + \lambda c_j^{\Delta}$, then the general condition of optimization will be the following:

$$z_j' - c_j(\lambda) = \sum_{i=1}^{m} \left(c_j^{t_0} + \lambda c_j^{\Delta}\right) \tau_{ij} - \left(c_j^{t_0} + \lambda c_j^{\Delta}\right) \leq 0, \quad j = 1,2,\ldots,n$$

Let's group those connections in the following way:

$$\sum_{i=1}^{m} c_j^{to}\tau_{ij} - c_j^{to} + \lambda\left(\sum_{i=1}^{m} c_j^{\Delta}\tau_{ij} - c_j^{\Delta}\right) \leq 0, \quad j = 1, 2, \ldots, n$$

and input the following notations: $\alpha_j = \sum_{i=1}^{m} c_j^{to}\tau_{ij} - c_j^{to}$, and $\beta_j = \left(\sum_{i=1}^{m} c_j^{\Delta}\tau_{ij} - c_j^{\Delta}\right)$. These constants are defined by the initial configuration: objective function coefficients and the corresponding basis solution, objective current coefficients with the condition that optimality hase not change during that period. In case of c_j^{to} initial value (when $\lambda = 0$) we have $\bar{x}^0$ optimization vertex and therefore, we get the following limitations: $\alpha_j \leq 0$.

The optimization variation does not take place in the $0 \leq \lambda \leq 1$ interval, therefore we also get the following:

$$\alpha_j + \lambda\beta_j \leq 0 :$$

In particular, when $\lambda = 1$ we get $\alpha_j + \beta_j \leq 0$.

The extreme condition will be written in the following general form: $\alpha_j + \lambda\beta_j \leq 0$, $j = 1, 2, \ldots, n$. Let's find the minimal value of λ at which at least one of these inequalities is violated for the first time.

Let's separate the negative and positive cases of β_j. λ restrictions will have the following form:

$$\lambda \geq -\alpha_j/\beta_j, \text{ for all } \beta_j < 0,$$
$$\lambda \leq -\alpha_j/\beta_j, \text{ for all } \beta_j > 0.$$

Let's input a new notetion:

$$\overline{\lambda} = \begin{cases} \min\limits_{\beta_j > 0} (-\alpha_j/\beta_j) \ , \\ +\infty, \quad \text{for all } \beta_j \leq 0 \ : \end{cases}$$

The optimal solution for all $\lambda = 0$ coincides with the optimal solution for all λ meeting the condition $0 \leq \lambda \leq \overline{\lambda}$. It is ensued that $\overline{\lambda}$ is the possible transition moment which is required. If $\overline{\lambda} = +\infty$, then there is no optimal plan variation. If $\overline{\lambda}$ is finite then it is necessary to consider two cases: the first one (a/) is the $\overline{\lambda}$ point, possible equivalent optimal plans and possible continuations in this case, and the second one (b/): if there is a new optimal plan and if the problem has no solution at $\lambda > \overline{\lambda}$.

a/ Assume that λ is finite, i.e. $\overline{\lambda} = -\alpha_k/\beta_k$ at k parameter corresponding value. It means that $z_k - c_k(\lambda) = 0$, from which follows that the optimization plan is not single. Actually, let's input k vector in the basis and according to the simplex method let's exclude one of the vectors from the previous basis. We will get a new optimal plan the objective value of which will stay unchaged. It follows more that by all null estimations and by all basis variations we can get many optimization equivalent vertices and all elements of their linear closure also have the same discussed optimization value.

b/ In this case, we consider the $\lambda > \bar{\lambda}$ case and the finite $\bar{\lambda}$ value. If above mentioned k vector coefficients all are not positive, i.e. $\tau_{ik} \leq 0$, then according to the simplex method, the objective function is becoms unlimited. This takes place any time when according to the increasing character of the objective function we will get the vector which is going to be included in the basis $z_k - c_k(\lambda) > 0$, but it becomes clear that the vector has no positive $\tau_{ik} > 0$ coordinate because of which we could exclude it from the basis. In this case, it is impossible to choose such $\theta > 0$ coefficient that any $x_i - \theta\tau_{ik} = 0$, where $i = \{1, ..., m\}$. Therefore, we get the optimization plan with $m + 1$ positive components i.e. the set of $\bar{a}_1, \bar{a}_2, ..., \bar{a}_m, \bar{a}_k$

linear-dependent vectors which corresponds to the nonangle vertex. Therefore, the linear objective function could not reach to its minimal value. This means that the hyperplane defined by linear function could not become the hyperplane of feasible polyhedron at any shift in the direction of gradient.

If any $\tau_{lk} > 0$, then $\bar{a}_k$ vector is included in the basis and another $\bar{a}_l$ vector is excluded from it. As we got the new basis by the simplex method then it corresponds to a new optimal solution, at that

$$\alpha'_j + \lambda\beta'_j \leq 0\,,\ j = 1, 2, \ldots, n (3)$$

inequalities are compatible.

Let's show that any $\lambda < \overline{\lambda}$ does not satisfy inequalities system (3). Really, for vector $\bar{a}_l$ excluded from the basis we will get the following:

$$\alpha'_l = -\alpha_k/\tau_{lk};\ \ \beta'_l = -\beta_k/\tau_{lk}, (4)$$

where $\tau_{lk} > 0$. Suppose that (3) takes place for any $\lambda < \overline{\lambda}$ then $\alpha'_j + \lambda\beta'_j \leq 0$, or according to (4) $-\alpha_k - \lambda\beta_k \leq 0$. As $\beta_k > 0$ then, from the latter inequality it follows that $\lambda \geq -\alpha_k/\beta_k = \overline{\lambda}$:

So, if $\overline{\lambda}$ is finite then in the case of the new basis it presents lower bound of variation.

4 Conclusion

The paper is devoted to the discussion of applied algorithms for data flows. The linear programming problems and solving simplex algorithm are considered. The problem is not about the simplex algorithm development, but the fact that the approaches processed in this sphere help when according to the problem assumption the coefficients of objective function vary in the result of the data flows analyses. We got that it is possible to introduce and develop the concepts and tools related to the simplex algorithm by approaches that solve flow linear optimization problems. The main result is the construction of the extrapolation mechanism that applies linear extrapolation and by predicting the stationary change of the flow explores and prepares possible succeeders of the optimality vertices in advance. This is important from the viewpoint of linear programming system optimization.

References

[1] L. Aslanyan, J. Castellanos, F. Mingo, H. Sahakyan, V. Ryazanov, Algorithms for Data Flows, International Journal \Information Theories and Applications}, ISSN 1310-05133, Volume 10, Number 3 (2003), pp. 279-282.

[2] Barbara G. Ryder and Marvin C. Paull, Incremental data-flow analysis algorithms, ACM Transactions on Programming Languages and Systems (TOPLAS), Volume 10 Issue 1 (1988), pp. 1-50.

[3] В. К. Леонтьев, Устойчивость в линейных дискретных задачах, "Проблемы Кибернетики", был. 35, 1979.

[4] A. Alaverdyan, Linear Programming With Changing Coefficients Of Objective Function, in \Transactions of the Institute for Informatics and Automation Problems of NAS RA}, Mathematical Problems of Computer Science, 2005, Yerevan.

[5] Ch. Papadimitriu, K. Steiglitz, Combinatorial optimization: Algorithms and Complexity, Prentice-Hall Inc, 1982.

[6] Ю. Кузнецов, В. Кузубов, А. Волощенко, Математическое программирование, Высшая школа, Москва, 1980.

[7] Б. Муртаф, Современное линейное Программирование, теория и практика, Москва, Мир, 1984.

Տվյալների հոսքերի վերլուծումը ըստ գծային ծրագրավորման մոդելի

Ա. Հ. Ալավերդյան, Վ. Մ. Կարախանյան, Վ. Ռ. Տոնոյան

Ամփոփում

Հոդվածը նվիրված է տվյալների հոսքերի ալգորիթմների կառուցման ընդհանուր խնդրի քննարկմանը մի կողմից, և որպես այս դասի մասնավոր խնդրի օրինակ՝ գծային ծրագրավորման խնդրի ուսումնասիրմանը, երբ հոսքային տվյալներից կախված փոփոխվում են նպատակային ֆունկցիայի գործակիցները: Խնդիրը ընթացիկ արդյունքի վերակառուցման մեջ է այնպիսի մի մոտեցմամբ, որը առավել պարզ է խնդրի լիակատար լուծումից ըստ հոսքի դինամիկ տվյալների բազմության: