Vasyl M. Teslyuk[1], **Yaroslav P. Kis**[2], **Tetiana M. Teslyuk**[3]

IMPROVING THE EFFICIENCY OF THE SOLUTION OF LINEAR PROGRAMMING TASKS APPLYING THE CUDA-TECHNOLOGY

The usage of a complex of methods for improving the efficiency of the solution of large dimensional linear programming tasks based on application of parallel data processing technology – CUDA has been offered. The specific features of the program and algorithmic peculiarities of implementing the abovementioned technology and the research outcomes are analysed.

Keywords: *large dimensional linear programming tasks, CUDA-technology, simplex method, Gauss-Jordan method for solving the systems of linear algebraic equations.*

Василь М. Теслюк, Ярослав П. Кісь, Тетяна М. Теслюк

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОЗВ'ЯЗАННЯ ЗАДАЧ ЛІНІЙНОГО ПРОГРАМУВАННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЇ CUDA*

У статті запропоновано використовувати комплекс методів для підвищення ефективності розв'язання задач лінійного програмування великої розмірності, що досягається застосуванням класичних підходів та технології паралельної обробки даних CUDA. Наведено специфіку програмної та алгоритмічної особливостей використання вищезазначеної технології та результати дослідження.

Ключові слова: *задачі лінійного програмування великої розмірності, технологія CUDA, симплекс метод, метод Гауса-Жордана для розв'язання системи лінійних алгебраїчних рівнянь.*

Форм. 2. Табл. 1. Рис. 3. Літ. 18.

Василий Н. Теслюк, Ярослав П. Кись, Татьяна Н. Теслюк

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ С ИССПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ CUDA

В статье предложено использовать комплекс методов для повышения эффективности решения задач линейного программирования большой размерности, что достигается применением классических подходов и технологии параллельной обработки данных CUDA. Описано специфику программной и алгоритмической особенностей использования вышеупомянутой технологии и результаты исследования.

Ключевые слова: *задачи линейного программирования большой размерности, технология CUDA, симплекс метод, метод Гаусса-Жордана для решения системы линейных алгебраических уравнений.*

Problem setting. It was estimated (Reklaitis, 1986) that nearly 75% of all practical optimization issues belong to linear programming tasks (LPT). These are the optimal resources allocation problems, transport costs tasks, network and scheduling, inventory control etc.

Globalization of the world markets, acute competition for shares at these markets, development of Internet technologies, the need to ensure high accuracy of opti-

[1] National University "Lviv Polytechnics", Ukraine.
[2] National University "Lviv Polytechnics", Ukraine.
[3] National University "Lviv Polytechnics", Ukraine.

* статтю підготовлено на основі доповіді на XII-му міжнародному науковому семінарі «Сучасні проблеми інформатики в управлінні, економіці, освіті та екології» (1–5 липня 2013 р., оз. Світязь – Київ).

mization results and the large variables number enhanced the search of design options for solving the linear programming tasks. Such optimization problems are of a large dimension where the variables number exceeds thousands.

Solution of linear programming tasks with large dimensions encounters some difficulties, namely, the need for huge personal computer resources. Accordingly, increase in efficiency of a wide dimension LPT solution which is achieved by the application of the classic approaches and parallel data processing technologies CUDA (Boreskov and Kharlamov, 2010; Sanders and Candrot, 2013) is the actual problem.

Latest research and publications analysis. A number of leading national and foreign scientists pay considerable attention to solving linear programming wide dimension problems, namely G.V. Dantzig (1949; 1960), B.G. Golshtejn and D.B. Yudin (1966), D.B. Yudin and B.G. Golshtejn (1969), M. Mesarovych, D. Mako and I. Takahara (1973), L. Lesdon (1975), V. Curkov (1981), R. Bellman (1960) and others.

The main idea of the proposed methods and approaches is based on the use of specific features of the double LPT and the decomposition approach, in which the original LPT of wide dimension has been split to easier ones and their solutions are found associated with the general solution of the original problem. G.V. Dantzig (1949; 1960) was the first to formulate the idea and to apply decomposition to LPT solving, which later was improved by B.G. Golshtejn and D.B. Yudin (1966), D.B. Yudin and B.G. Golshtejn (1969), M. Mesarovych et al. (1973), L. Lesdon (1975), V. Tsurkov (1981) etc. Additional attention should be paid to the application of a Bellman's dynamic programming method (Bellman, 1960; Cormen et al., 2001).

Charcteristics of decomposition methods, in most cases, relate to the limitation matrix breaking into smaller dimensions matrices approach. However, unfortunately, the majority of common decomposition methods do not give an opportunity to use parallel algorithms for solving linear large dimension programming tasks.

The research objective is the analysis of methods and tools of solving linear programming tasks of wide dimension and the related CUDA technology application.

Unresolved issues. The choice of effective methods and tools for linear programming problems of high dimensionality solution with the use of parallel data processing technology remains an unresolved issues.

Key research findings. The majority of optimization problems of economic nature refers to the linear programming group. The standard formulation of such LPT is done in the following way: variables values $x_1, x_2, \ldots, x_n$, that meet the following equations (restrictions) system are found (Reklaitis et al., 1986):

$$\begin{aligned} a_{11}x_1 + \quad & a_{12}x_2 + \quad \ldots \quad + a_{1n}x_n \quad = b_1, \\ a_{21}x_1 + \quad & a_{22}x_2 + \quad \ldots \quad + a_{2n}x_n \quad = b_2, \\ \ldots \quad & \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + \quad & a_{m2}x_2 + \quad \ldots \quad a_{mn}x_n \quad = b_m \end{aligned} \tag{1}$$

the smallest value of objective function is given:

$$\min F = c_1x_1 + c_2x_2 + \ldots + c_nx_n, \tag{2}$$

where c_i are the constants; a_{ji} – the restriction matrix coefficients; b_j – free terms column; n – variables number, m – constraints number.

In general case, for solving the tasks (1–2) the simplex-method or its improvements are used. Its algorithm includes the following steps (Reklaitis et al., 1986):

Step 1. The initial admissible base solution should be determined by equating to zero non-base n-m variables with the use of the standard form linear model of LPT.

Step 2. The variable, which will be included in the new basis and the value increase/decrease of which will provide improvements of the objective function value, should be selected from the number of the current non-base (equal to zero) variables. If such variable is absent, the calculations should be stopped as the current base, meaning that the solution is optimal. Otherwise, go to step 3.

Step 3. The variable which appears in the list of basic variables and must take a null value (become non-base) at the entry to the basis of the new variable should be chosen from the number of the current basis variables.

Step 4. A new base of the task solution that corresponds to the new list of non-base and basic variables should be found. Proceed to step 2.

In the case of large dimension linear programming (1–2) the amount of computations can be significantly reduced by solving the problem of double task (Reklaitis et al., 1986), which is possible when the number of constraints in the double task is less.

Another group, widely used for solving linear programming wide dimension tasks is, as mentioned above, based on the decomposition approach (Lesdon, 1975; Tsurkov, 1981).

In accordance with the algorithm described above, the largest volume of calculations is performed in step 4, where a new base task solution should be found.

However, it is known that during the execution of step 4, the Gauss-Jordan scheme (Gregori and Miller, 2013; Demydovich and Maron, 1966) is used. Thus, it is reasonable to use CUDA technology for solving this problem in order to increase the effectiveness of linear programming tasks solution.

Since the Gauss-Jordan scheme algorithm involves operations with matrices, exactly this part should be performed in parallels.

In step 4 of the simplex-method algorithm the two types of operations (Taha, 1985) are involved. The first type is associated with the calculations of the basic values, and the second one with all other lines, including the target function.

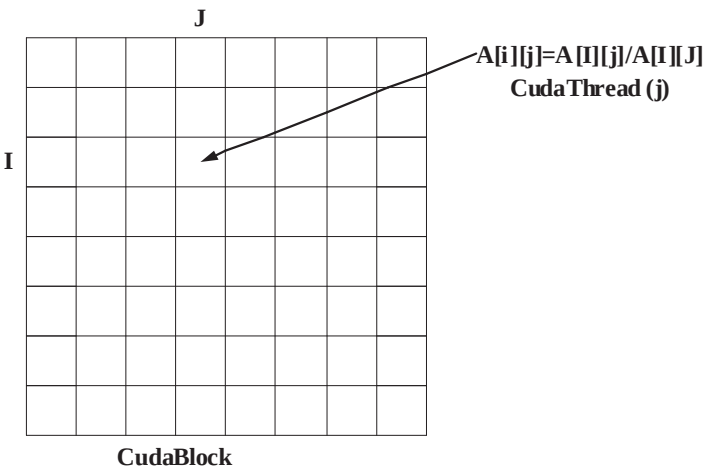
During the first type of operations implementation, the value of each base tape element is divided on the value of the base element. In the process of CUDA technology application the given operation will occur on the graphic adapter, copies of simplex table data will be also used, and the scheme of this operation type is shown below (Figure 1). Thus, with the use of this scheme the parallel operation of all leading tape elements values division on the leading elements values is implemented.

The second type of operations are designed to define the new values in a simplex table, which is determined by multiplication of the related leading tape element value on the leading column coefficient of the previous level and the obtained result subtraction from the previous value of the simplex table element.

The implementation scheme of this operation type with the use of CUDA technology is shown in Figure 2.

The analysis of the research outcomes. In the process of CUDA technology application for solving large dimension LPT the NVidia GTX650 video card has been

used. The comparative results of time spent on one iteration during the solution of the LPT with the use of a computer, AMD Athlon 64 x 23800 + 2Ghz microprocessor and NVidia GTX650 video card are shown in Table 1 and Figure 3.



Note: I – number of the leading line, J – number of the leading column.

Figure 1. **Operating scheme of the algorithm core for the first type of operations realization with the use of CUDA,** *developed by the authors*

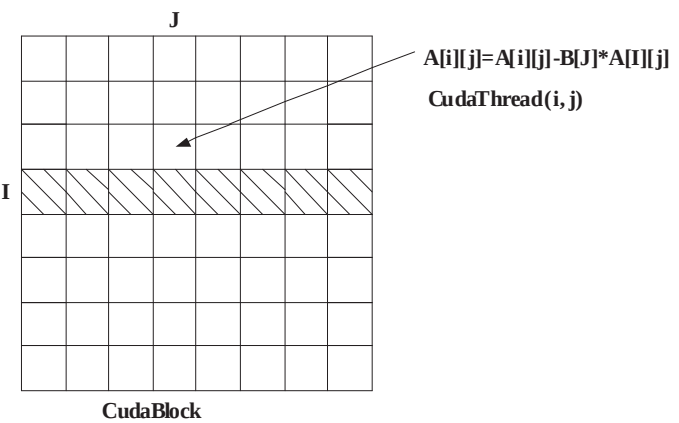


Figure 2. **The second type of operation implementation scheme with the use of CUDA,** *developed by the authors*

The obtained data demonstrate that for a small dimension simplex table of 100–150 variables, the time of simplex-table items transfer with the use of NVidia GTX650 video card is larger. Such results can be explained by the fact that the total time includes forwarding data from RAM to the video card memory, which has crucial influence on the given dimensions.

For the larger dimensions of the simplex table the advantage from the CUDA technology usage only grows with the increasing number of parameters that can reach tens or even hundreds of times.

Table 1. **Comparison of the time spent on one iteration in the process of LPT solving,** *developed by the authors*

The size of the coefficients matrix of the system of linear algebraic equations (SLAE)\execution time in milliseconds	*CPU AMD Athlon64x2 3800+ 2Ghz*	*GPU NVidia GTX650*
64x64	3	27
128x128	25	64
256x256	243	140
512x512	1984	368
1024x1024	14498	1224
2048x2048	121052	7812

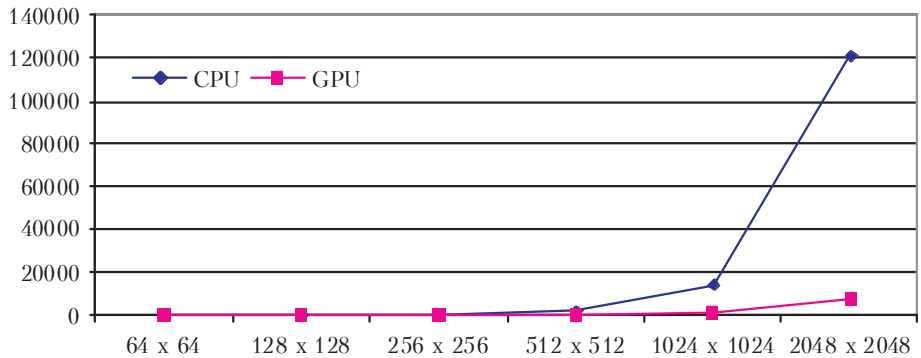


Figure 3. **Dependence of the Gauss-Jordan algorithm execution time on matrices size (CPU and GPU),** *developed by the authors*

Conclusions:

1. Application of the parallel data processing technology with the use of CUDA technology for solving the linear programming tasks allows to reduce the execution time significantly.

2. Increase in the video card cores number allows to reduce the time for the LPT solving. However, the speed of the data transfer through the computer data line may slow the process down. Therefore, it is advisable to use the methods of direct video card memory access, buffering or previously downloading the data in the video card memory.

3. Due to the absence of data processing means with double precision, the calculations error should be controlled in the video card and, if necessary, special approaches, methods and algorithms to solve this shortage should be used.

References:

Беллман Р. Динамическое программирование. – М.: Изд-во иностранной литературы, 1960. – 400 с.

Боресков А.В., Харламов А.А. Основы работы с технологией CUDA. – М.: ДМК Пресс, 2010. – 232 с.

Вечканов Г.С. Экономическая теория: Учебник для вузов. – 3-е изд. – СПб.: Питер, 2011. – 512 с.

Гольштейн Е.Г., Юдин Д.Б. Новые направления в линейном программировании. – М.: Советское радио, 1966. – 524 с.

Грегори К., Миллер Э. C++ AMP: построение массивнопараллельных программ с помощью Microsoft Visual C++. – М.: ДМК Пресс, 2013. – 412 с.

Демидович Б.П., Марон И.А. Основы вычислительной математики. – 3-е изд. – М.: Наука, 1966. – 664 с.

Київський національний університет імені Тараса Шевченка: Незабутні постаті / Авт.-упор. О. Матвійчук, Н. Струк; Ред. кол.: В.В. Скопенко, О.В. Третяк, Л.В. Губерський, О.К. Закусило, В.І. Андрейцев, В.Ф. Колесник, В.В. Різун та ін. – К.: Світ успіху, 2005. – 464 с.

Лэсдон Л.С. Оптимизация больших систем. – М.: Наука, 1975. – 432 с.

Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем / Пер. с англ. И.Ф. Шахнова. – М.: Мир, 1973. – 344 с.

Полтерович В.М. Теория оптимального распределения ресурсов Л.В. Канторовича в истории экономической мысли // Журнал Новой экономической ассоциации: 2012.– №1. – С. 176–180.

Реклейтис Г., Рейвиндрон А., Рэгсдел К. Оптимизация в технике: В 2-х кн. / Пер. с англ. – М.: Мир, 1986. – Кн. 2. – 320 с.

Сандерс Дж., Кэндрот Э. Технология CUDA в примерах: введение в программирование графических процессоров. – М.: ДМК Пресс, 2013. – 232 с.

Таха Х. Введение в исследование операций: В 2-х кн. / Пер. с англ. – М.: Мир, 1985. – Кн. 1. – 479 с.

Цурков В.И. Декомпозиция в задачах большой размерности. – М.: Наука, 1981. – 352 с.

Юдин Д.Б., Гольштейн Е.Г. Линейное программирование. – М.: Наука, 1969. – 424 с.

Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. (2001). Introduction to Algorithms. 2nd ed. MIT Press & McGraw-Hill.

Dantzig, G.B. (1949). Programming in linear structure. Econometrica, 17: 73–74.

Dantzig, G.B. (1960). Decomposition Principle for Linear Programs. Operations Research, 8: 101–111.

Стаття надійшла до редакції 23.07.2013.