

## К АНАЛИЗУ МЕТОДА ИМИТАЦИИ ОТЖИГА В МНОГОЭКСТРЕМАЛЬНОМ СЛУЧАЕ\*

*С. М. Ермаков, Д. В. Куликов, С. Н. Леора*

Санкт-Петербургский государственный университет,  
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Существует ряд важных прикладных задач, где число подлежащих вычислению глобальных экстремумов велико или даже бесконечно велико. К числу таких задач относятся, например, некоторые задачи планирования эксперимента, задачи о решении больших систем уравнений и ряд других. Если речь идет об одном экстремуме функции многих переменных, наиболее распространенным численным методом является метод имитации отжига, который также успешно применяется в дискретных задачах большого объема (задача о коммивояжере). В дискретных задачах известно, что метод имитации отжига ищет равные глобальные экстремумы с равной вероятностью. Непрерывный случай до сих пор не был исследован. Предполагалось, что равные экстремумы нужно находить последовательно, разделяя их окрестности в процессе вычислений. Это не всегда эффективный метод, особенно в случае, когда множество экстремумов заполняет некую область в  $\mathbf{R}^n$ .

Полученные в данной статье результаты намечают общий подход к проблеме. Приводятся вычислительные примеры, свидетельствующие об эффективности подхода. На его базе возможно создание программ алгоритмов, указывающих локализацию корней больших систем уравнений. Также можно отметить, что очень многие задачи планирования регрессионных экспериментов имеют бесконечное множество решений. Библиогр. 4 назв. Ил. 5.

*Ключевые слова:* имитация отжига, глобальный экстремум, планирование эксперимента.

Метод имитации отжига [1–3] широко применяется для решения прикладных задач по поиску абсолютного экстремума. Его математическая сторона кратко может быть изложена следующим образом. Пусть  $f(X)$ ,  $X \in D$ ,  $D \in R^s$ , ограниченная функция,  $|f(X)| \leq M$ . Как правило,  $D$  также ограничена в  $R^s$ . Случай неограниченной  $D$  требует отдельного рассмотрения. Пусть также  $T$  – вещественный параметр,  $T \in R^1$ . Можно показать, что функция (плотность распределения)

$$\overline{\mathfrak{M}}(T, X) = C(T) \exp\left(-\frac{f(X)}{T}\right), \quad (1)$$

где  $C$  – константа нормировки, зависящая от  $T$ , при  $T \rightarrow 0$  слабо сходится к  $\delta$ -функции, сосредоточенной в точке глобального минимума  $f(X)$ .

Алгоритм состоит в моделировании методом Метрополиса [1] последовательности плотностей (1) с убывающими значениями  $T_1 > T_2 > \dots > T_n$ . Реализации соответствующих случайных величин, полученные при  $T_k$ , служат исходными для алгоритма Метрополиса при  $T_{k+1}$ ,  $k = 1, 2, \dots, n - 1$ .

В [4] предлагалось для нахождения глобального максимума использовать другую функцию типа  $\overline{\mathfrak{M}}(T, X)$ , а именно

$$\mathfrak{M}(X, n) = \frac{f^n(X)}{\int_D f^n(X) dX}. \quad (2)$$

\*Работа выполнена при финансовой поддержке РФФИ (грант № 14-01-00271-а).

© Санкт-Петербургский государственный университет, 2017

При  $n \rightarrow \infty$  функция  $\mathfrak{M}(X, n)$  также сходится к функции глобального максимума. Легко показать, что применение алгоритма Метрополиса приводит к модификации метода имитации отжига, где параметр  $n$  играет роль  $T$ . Далее мы будем рассматривать алгоритмы, основанные на моделировании  $\mathfrak{M}(X, n)$  при  $n \rightarrow \infty$ . При этом в работе специально рассматривается случай двух и более одинаковых максимумов (минимумов) исследуемой функции.

Для конкретности изложения предварительно укажем две задачи, где требуется вычислять глобальный экстремум.

**Задача 1.** Требуется найти приближенные значения вещественных корней системы нелинейных уравнений  $g_l(X) = 0, X(x_1, \dots, x_s), l = 1, 2, \dots, s$ .

Задача будет решена, если найти все точки глобального минимума функции

$$H(X) = \sum_{l=1}^s p_l g_l^2(X) = \min, \quad (3)$$

$p_l$  — заданные константы,  $p_l > 0$ . (Очевидно задачу легко переформулировать в задачу нахождения максимума.)

**Задача 2.** Построение точного  $D$ -оптимального плана регрессионного эксперимента. Самая простая из этих задач в случае двух переменных (построение плоскости по трем точкам  $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ ) требует нахождения наибольшего значения определителя

$$\det \|1, x_i, y_i\|_{i=1}^3. \quad (4)$$

Если область, где строится оптимальный план, является квадратом  $[0, 1]^2$ , решение известно. Это  $(0, 0)$ ,  $(1, 0)$  и любая точка отрезка  $x \in [0, 1]$  при  $y = 1$ . Также оптимальными будут планы, полученные из этого плана поворотами на  $\frac{\pi r}{2}, r = 1, 2, 3$ .

В общем случае приходится иметь дело с максимизацией больших определителей такого типа, где явное решение найти невозможно.

Отличительной особенностью двух упомянутых задач является существование большого (может быть бесконечного, как в задаче 2) числа равных глобальных экстремумов. Если случай единственного экстремума изучен в (1) достаточно полно, то случай множественных экстремумов, особенно для (2) изучен недостаточно. Ниже изучаются некоторые свойства и особенности алгоритмов, когда (2) является плотностью по отношению к заданной мере и единственность экстремума не предполагается.

Пусть  $(\mathfrak{A}, D, \mu)$  — пространство с  $\sigma$ -алгеброй подмножеств  $D$  и конечной мерой  $\mu$ ,  $f(X) \in L(D, \mu)$ ,  $f$  является  $\mu$ -интегрируемой, ограниченной, неотрицательной функцией в  $D$ , имеющей некоторое множество  $M$  глобальных максимумов.

Рассмотрим последовательность распределений, заданных следующим образом:

$$\mathcal{P}_n(\mathcal{A}) = \frac{\int_{\mathcal{A}} f^n(x) \mu(dx)}{\int_D f^n(x) \mu(dx)}, \quad \forall \mathcal{A} \in \mathfrak{A}, \quad n = 1, 2, \dots \quad (5)$$

Вопрос о характере предельного распределения в общем случае не является простым. Мы остановимся на двух специальных случаях, которые необходимы для конструирования алгоритма.

**I.** Пусть  $\mu$  — дискретная мера, сосредоточенная в точках  $x_1, x_2, \dots$ . Предварительно докажем следующую лемму.

**Лемма.** Пусть  $q_i$  — вещественные числа,  $0 \leq q_i < 1$  и

$$\sum_{i=1}^{\infty} q_i < \infty, \quad (6)$$

тогда справедливо равенство

$$\lim_{k \rightarrow \infty} \sum_{i=1}^{\infty} q_i^k = 0. \quad (7)$$

**Доказательство.** Для  $\forall \varepsilon > 0$  в силу условия (6) найдется  $N \in \mathbb{N}$ , такое что  $\sum_{i=N+1}^{\infty} q_i < \varepsilon/3$  и, поскольку  $q_i < 1$ , такое  $k \in \mathbb{N}$ , что для любого  $i < N$  верно  $q_i^k < 2^{-i} \cdot \varepsilon/3$ .

Тогда для выбранных  $N$  и  $k$  будем иметь

$$\sum_{i=1}^{\infty} q_i^k = \sum_{i=1}^N q_i^k + \sum_{i=N+1}^{\infty} q_i^k \leq \frac{\varepsilon}{3} \sum_{i=1}^N 2^{-i} + \frac{\varepsilon}{3} < \varepsilon, \quad \text{т. к.} \quad \sum_{i=N+1}^{\infty} q_i^k < \sum_{i=N+1}^{\infty} q_i. \blacksquare$$

Пусть теперь  $Q$  — дискретное мультимодальное распределение с  $m$  одинаковыми модами

$$Q_k = \begin{pmatrix} x_1, & x_2, & \dots \\ c_k q_1^k, & c_k q_2^k, & \dots \end{pmatrix}, \quad k = 1, 2, \dots,$$

причем  $x_{i_1}, x_{i_2}, \dots, x_{i_m}$  — одинаковые моды,  $q_{i_1} = q_{i_2} = \dots = q^*$ , а  $c_k$  — константа нормировки,  $c_k = (\sum_{i=1}^m q_i^k)^{-1}$ . Обозначим

$$Q_0 = \begin{pmatrix} x_{i_1}, & x_{i_2}, & \dots, & x_{i_m} \\ 1/m & 1/m & \dots & 1/m \end{pmatrix}.$$

**Теорема 1.**  $Q_k$  при  $k \rightarrow \infty$  слабо сходится к  $Q_0$ .

**Доказательство.** Покажем, что если  $q_j \notin \{q_{i_1}, \dots, q_{i_m}\}$ , то  $c_k q_j^k \rightarrow 0$  при  $k \rightarrow \infty$ :

$$\frac{q_j^k}{\sum_{i=1}^m q_i^k} = \frac{(q_j/q^*)^k}{\sum_{j=1}^m (q_j/q^*)^j} = \frac{(q_j/q^*)^k}{m + \sum_{j \notin \{i_1, \dots, i_m\}} (q_j/q^*)^k}.$$

При  $k \rightarrow \infty$  второе слагаемое знаменателя стремится к нулю по лемме и очевидно вся дробь стремится к нулю также.  $\blacksquare$

Таким образом, предельным распределением является равномерное распределение на множестве равных экстремумов.

**II.** Вернемся к общему случаю. Обозначим  $G = \max(f(X))$ ,  $X \in D$ , и положим

$$M = \{x \in D : f(X) = G\}. \quad (8)$$

Не умаляя общности, будем считать, что  $G = 1$ .

Справедлива следующая теорема.

**Теорема 2.** Если  $\mu(M) > 0$ , то справедливо равенство

$$\lim_{n \rightarrow \infty} \frac{\int_A f^n(x) \mu(dx)}{\int_D f^n(x) \mu(dx)} = \frac{\mu(A \cap M)}{\mu(M)}. \quad (9)$$

ДОКАЗАТЕЛЬСТВО. Как и в дискретном случае имеем

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\int_{\mathcal{A}} f^n(x) \mu(dx)}{\int_D f^n(x) \mu(dx)} &= \lim_{n \rightarrow \infty} \frac{\int_{\mathcal{A} \cap M} f^n(x) \mu(dx) + \int_{\mathcal{A} \setminus M} f^n(x) \mu(dx)}{\int_M f^n(x) \mu(dx) + \int_{D \setminus M} f^n(x) \mu(dx)} = \\ &= \lim_{n \rightarrow \infty} \frac{\mu(\mathcal{A} \cap M) + \int_{\mathcal{A} \setminus M} f^n(x) \mu(dx)}{\mu(M) + \int_{D \setminus M} f^n(x) \mu(dx)} = \frac{\mu(\mathcal{A} \cap M)}{\mu(M)}, \end{aligned}$$

поскольку интеграл на  $D \setminus M$  стремится к нулю. Напомним, что на множествах  $\mathcal{A} \setminus M$  и  $D \setminus M$  выполнено  $f(x) < 1$  для всех  $x \pmod{\mu}$ . ■

Случай  $\mu(M) = 0$  теоретически интересен, но требует отдельных исследований и дополнительных предположений относительно поведения  $f(X)$  в окрестности точек экстремума.

Тем не менее, для построения прикладных алгоритмов (решения задач 1 и 2, например) результат теоремы 2 может быть весьма полезным.

Введем в рассмотрение функцию ( $0 < \varepsilon < 1$ )

$$G_\varepsilon(X) = \begin{cases} F(X), & F(X) \leq 1 - \varepsilon, \\ 1 - \varepsilon, & F(X) > 1 - \varepsilon. \end{cases} \quad (10)$$

Если  $\mu(\{X : F(X) = 1\}) = 0$ , мы будем предполагать, что при каждом  $\varepsilon > 0$  справедливо неравенство  $\mu(\{X : G_\varepsilon(X) = 1 - \varepsilon\}) > 0$ .

Во всяком случае так будет в случае меры Лебега  $\mu$  для непрерывной  $F$ . При этом множество  $\{X : G_\varepsilon(X) = 1 - \varepsilon\}$  может состоять из набора изолированных, при достаточно малых  $\varepsilon$ , множеств  $M_i, i = 1, 2, \dots$ , на каждом из которых  $G_\varepsilon(X) = \text{const}$ , но  $\mu(M_i)$  не обязано, вообще говоря, совпадать с  $\mu(M_j)$  при  $i \neq j$ .

В соответствии с теоремой 2 абсолютные экстремумы  $G_\varepsilon(X)$  будут распределены равномерно на множестве  $M_\varepsilon = \{X : G_\varepsilon(X) = 1 - \varepsilon\}$ .

Алгоритм, являющийся аналогом метода имитации отжига, предполагает задание последовательности  $(\varepsilon_k, n_k), k = 1, 2, \dots$  При каждом  $\varepsilon_k$  и  $n_k$  метод Метрополиса используется для моделирования  $(G_{\varepsilon_k}(X))^{n_k} C_k$ . Здесь  $\varepsilon_k$  — убывающая, а  $n_k$  — возрастающая последовательности и  $C_k$  — константа нормировки. При каждом  $k$  заполняется массив  $m$  реализаций моделируемых случайных величин.

Здесь, очевидно, возможны различные усовершенствования, зависящие от характера задачи. Более сложная, но во многом аналогичная ситуация возникает при решении задачи 2.

**Численный эксперимент.** Ниже приводятся простейшие иллюстративные примеры использования алгоритмов, основанных на выше изложенных теоретических результатах.

**Пример 1.** Рассмотрим функцию, определенную на промежутке  $[0; 2\pi]$  и имеющую два одинаковых максимума:

$$f(x) = \begin{cases} \sin(x), & 0 \leq x \leq \pi, \\ \sin(x - \pi), & \pi < x \leq 2\pi. \end{cases}$$

Введем функцию согласно (10):

$$f_\varepsilon(x) = \begin{cases} f(x), & f(x) \leq 1 - \varepsilon, \\ 1 - \varepsilon, & f(x) > 1 - \varepsilon. \end{cases}$$

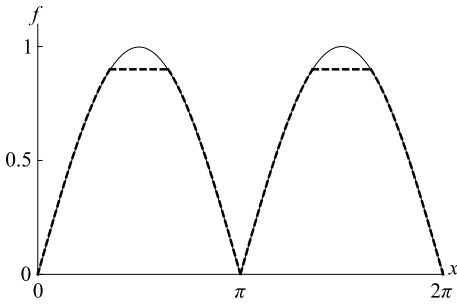


Рис. 1. Графики функций:  $f(x)$  — сплошная линия,  $f_\varepsilon(x)$  — пунктирная линия.

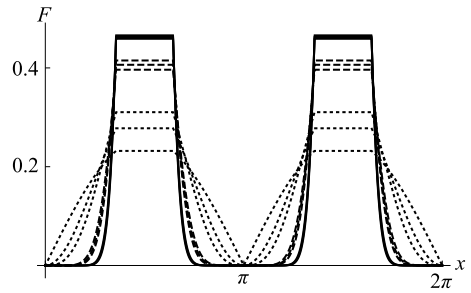


Рис. 2. Графики функций  $F(n, f(x))$ , значениям  $n = 1, 2, 3$  соответствуют пунктирные линии,  $n = 8, 9, 10$  — штриховые и  $n = 18, 19, 20$  — сплошные.

Пусть для определенности  $\varepsilon = 0.1$ . Графики функций  $f(x)$  и  $f_\varepsilon(x)$  изображены на рис. 1. Функция имеет два глобальных максимума в точках  $\pi/2$  и  $3\pi/2$ .

Рассмотрим класс функций  $F(n, f(x))$ , где  $n$  — параметр,  $f(x)$  — функция, глобальные экстремумы которой подлежат определению

$$F(n, f(x)) = \frac{f_\varepsilon^n(x)}{\int_0^{2\pi} f_\varepsilon^n(x) dx}. \quad (11)$$

Функция вида (11) при  $n \rightarrow \infty$ , как было доказано выше, стремится к равномерному распределению на множестве точек одинаковых глобальных экстремумов. Построенные для ряда значений  $n$  графики, приведенные на рис. 2, наглядно это демонстрируют. В соответствии с теоремой 2 абсолютные экстремумы будут распределены на множестве  $M_\varepsilon = \{x : f_\varepsilon(x) = 1 - \varepsilon\}$ . В данном случае  $M_\varepsilon = M_1 \cup M_2$ , причем  $\mu(M_1) = \mu(M_2)$ .

**Пример 2.** Рассмотрим функцию, определенную на промежутке  $[0; 5\pi/4]$  и имеющую два одинаковых максимума, отличающихся «крутизной» вершин:

$$g(x) = \begin{cases} \sin(x), & 0 \leq x \leq \pi, \\ \sin 4(x - \pi), & \pi < x \leq 5\pi/4. \end{cases}$$

В этом случае для поиска экстремумов можно рассматривать класс функций вида

$$G(n, g(x)) = \frac{g_\varepsilon^n(x)}{\int_0^{5\pi/4} g_\varepsilon^n(x) dx}.$$

Построенные для ряда значений  $n$  графики функции  $G(n, f(x))$ , приведены на рис. 3.

В данном случае абсолютные экстремумы будут распределены равномерно также на двух изолированных множествах  $M_1$  и  $M_2$ ,  $M_\varepsilon = M_1 \cup M_2$ , причем  $\mu(M_1) \neq \mu(M_2)$ .

**Пример 3.** Рассмотрим функцию двух переменных

$$z(x, y) = \frac{1}{1 + 5(y - x^2)^2 + 2(y + x^2 - 1)^2}.$$

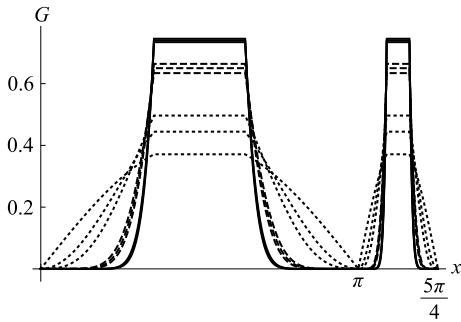


Рис. 3. Графики функций  $G(n, f(x))$ , значениям  $n = 1, 2, 3$  соответствуют пунктирные линии,  $n = 8, 9, 10$  — штриховые и  $n = 18, 19, 20$  — сплошные.

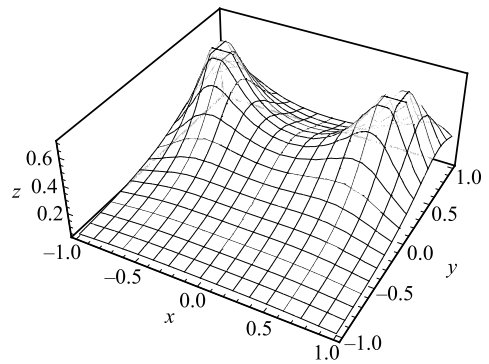


Рис. 4. График функции  $Z(1, z(x, y))$ .

Известно, что данная функция имеет два одинаковых максимума в точках  $(x_{\max}, y_{\max}) = (\pm 1/\sqrt{2}, 1/2)$ .

Пусть  $D = \{(x, y) : -1 \leq x \leq 1, -1 \leq y \leq 1\}$ . Выберем  $\varepsilon = 0.1$ . Моделирование проводится аналогично двумерному случаю. Рассматривается класс функций  $Z(n, z(x, y))$ , введенных согласно (10). На рис. 4 приведен график функции  $Z(1, z(x, y))$ .

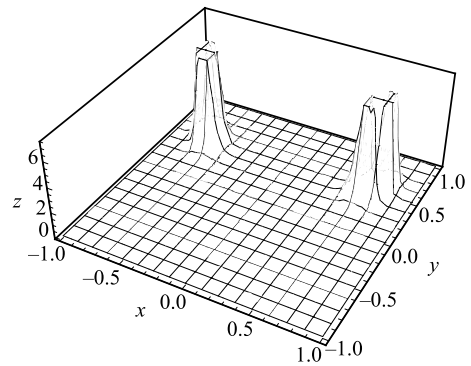
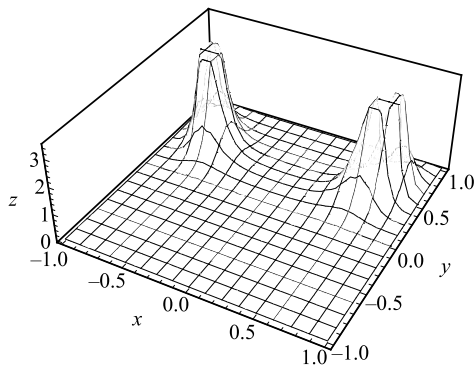


Рис. 5. Графики функции  $Z(n, z(x, y))$  при  $n = 5$  (слева),  $n = 15$ .

На рис. 5 приведены графики функции  $Z(n, z(x, y))$  при различных значениях  $n$ . Абсолютные экстремумы распределены равномерно на двух изолированных множествах.

### Литература

1. Metropolis V. N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E. Equations of state calculations by fast computing machines // J. Chem. Phys. 1953. Vol. 21. P. 1087–1092.
2. Ingber L. Adaptive simulated annealing (ASA): Lessons learned // J. Control and Cybernetics. 1995.

3. Winkler G. Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction. Vol. 27. Springer Science and Business Media, 2012.

4. Ермаков С. М. Метод Монте-Карло и смежные вопросы. М.: Наука, 1975.

Статья поступила в редакцию 26 сентября 2016 г.; рекомендована в печать 22 декабря 2016 г.

#### Сведения об авторах

Ермаков Сергей Михайлович — доктор физико-математических наук, профессор;  
sergej.ermakov@gmail.com

Куликов Денис Валерьевич — студент; altingfest@gmail.com

Леора Светлана Николаевна — кандидат физико-математических наук, доцент; leora2008@mail.ru

#### TOWARDS THE ANALYSIS OF THE SIMULATED ANNEALING METHOD IN THE MULTIEXTREMAL CASE

Sergey M. Ermakov, Denis V. Kulikov, Svetlana N. Leora

St. Petersburg State University, Universitetskaya nab., 7–9, St. Petersburg, 199034, Russian Federation;  
sergej.ermakov@gmail.com, altingfest@gmail.com, leora2008@mail.ru

There are many applications, where the number to be calculated global extrema is large, or even infinite. These problems include, for example, some experimental design problems, the problem of solving large systems of equations, and several others. If this is single extreme of a function of several variables, one of the commonly used numerical algorithms is the Simulated Annealing, which is also successfully used in high volume discrete problems (traveling salesman problem). In discrete problems known, that the method of simulated annealing searches equal global extremes with equal probability. Continuous case still has not been investigated. It was assumed that equal extremes to be found consistently, sharing their neighborhood during the computation. It is not always effective method, especially in the case when multiple extrema fills a certain region in  $\mathbf{R}^n$ .

The results obtained in this article the general outline of the approach to the problem. We give computational examples showing the effectiveness of the approach. It can be used to create programs, algorithms, indicating the localization of the roots of large systems of equations. It can also be noted, that many planning tasks regression experiments are an infinite number of solutions. Refs 4. Figs 5.

*Keywords:* simulated annealing, global extremum, experimental designing.

#### References

1. Metropolis V. N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., “Equations of state calculations by fast computing machines”, *J. Chem. Phys.* **21**, 1087–1092 (1953).
2. Ingber L., “Adaptive simulated annealing (ASA): Lessons learned”, *J. Control and Cybernetics* (1995).
3. Winkler G., *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction* **27** (Springer Science and Business Media, 2012).
4. Ermakov S. M., *The Monte Carlo Method and Related Questions* (Nauka, Moscow, 1975) [in Russian].

**Для цитирования:** Ермаков С. М., Куликов Д. В., Леора С. Н. К анализу метода имитации отжига в многоэкстремальном случае // Вестник СПбГУ. Математика. Механика. Астрономия. 2017. Т. 4 (62). Вып. 2. С. 220–226. DOI: 10.21638/11701/spbu01.2017.205

**For citation:** Ermakov S. M., Kulikov D. V., Leora S. N. Towards the analysis of the simulated annealing method in the multiextremal case. *Vestnik SPbSU. Mathematics. Mechanics. Astronomy*, 2017, vol. 4 (62), issue 2, pp. 220–226. DOI: 10.21638/11701/spbu01.2017.205