ИДЕНТИФИКАЦИЯ АВТОРЕГРЕССИОННОГО ПРОЦЕССА В СРЕДЕ MATLAB

DOI: 10.36724/2072-8735-2021-15-7-34-38

Шерстнева Алина Анатольевна,
Сибирский Государственный Университет Телекоммуникаций и Информатики, Новосибирск, Россия, **shers7neva@gmail.com**

Шерстнева Ольга Григорьевна,
Сибирский Государственный Университет Телекоммуникаций и Информатики, Новосибирск, Россия, **sherstneva@ngs.ru**

Manuscript received 18 February 2021;
Accepted 12 April 2021

Ключевые слова: *регрессионный анализ, метод наименьших квадратов, прогнозирование, изменение данных, идентификация, авторегрессионный процесс*

Целью статьи является применение методов регрессионного анализа в задаче оценки показателей надежности инфокоммуникационной системы и прогнозирования тренда данных. Задача расчета показателей надежности инфокоммуникационных систем основывается, как правило, на статистических данных, сбор и обработку которых осуществляет система мониторинга. Для получения максимально приближенных к реальным практическим результатам расчётных показателей необходимо проводить большое число измерений. В этом смысле теория фильтрации находит широкое применение в разнообразных задачах оценивания. Популярность теории обусловлена возможностью эффективного решения технических вопросов и реализацией через программы математического моделирования. Статья направлена на рассмотрение вопросов прогнозирования тренда данных для расчета параметров инфокоммуникационных систем. Одним из эффективных путей решения является применение методов регрессионного анализа. Выполнена разработка и исследование приведенных в работе математических и программных моделей. На основании теоретических выкладок происходит определение методов решения задачи и проведение экспериментальных исследований. В статье задача идентификации авторегрессионного процесса решена через систему уравнений Юла-Уокера. Помимо теоретических выкладок, разработана программа позволяющая автоматизировать процесс вычислений в среде Matlab. В результате предлагается сравнение результатов идентификации интересующего временного ряда переменных через систему Юла-Уокера с классической оценкой параметров методом наименьших квадратов. Результаты приведены в графическом виде.

Информация об авторах:

Шерстнева Алина Анатольевна, *к.т.н., Сибирский Государственный Университет Телекоммуникаций и Информатики, Новосибирск, Россия*

Шерстнева Ольга Григорьевна, *к.т.н., Сибирский Государственный Университет Телекоммуникаций и Информатики, Новосибирск, Россия*

Для цитирования:
Шерстнева А.А., Шерстнева О.Г. Идентификация авторегрессионного процесса в среде Matlab // T-Comm: Телекоммуникации и транспорт. 2021. Том 15. №7. С. 34-38.

For citation:
Sherstneva A.A., Sherstneva O.G. (2021) Identification of an autoregressive process using Matlab. *T-Comm*, vol. 15, no. 7, pp. 34-38. (*in Russian*)

Введение

При исследовании процессов функционирования инфокоммуникационных систем и сетей связи, как правило, используют определенный алгоритм действий, состоящий из нескольких этапов. Первоначально, инфокоммуникационная система или сеть связи раскладывается на отдельные типовые элементы (объекты), функциональные блоки или используемые линии связи. Затем рассматривается внутритехнологический процесс их эксплуатации с учетом надежности функционирования. Далее составляется математическая модель в виде графа состояний. Число состояний зависит от целей исследования и дополнительных не только внешних, но и внутренних характеристик исследуемого объекта. Примером составления математической модели может служить граф состояний, опубликованный в [1]. Результатом исследования графа состояний является получение расчетных формул для определения ключевых показателей функционирования системы. Компонентами определяющих выражений являются переменные величины, к которым относятся вероятности разного типа отказов объекта, обнаруживаемые системой мониторинга, интенсивность восстановления, а также ошибки I, II рода системы контроля работоспособного состояния объекта.

Интерес представляет проблема прогнозирования этих данных при изменении условий эксплуатации объекта. При решении задачи прогнозирования тренда данных, классическим методом является оценка параметров временного ряда и дисперсии методом наименьших квадратов. Рассматриваемая задача решается путем составления модели многомерной регрессии для оценки прогнозируемых параметров, фундаментальные исследования которой приводят к оценке параметров авторегрессионного процесса [2-4].

Постановка задачи

Процесс авторегрессии p-го порядка описывается уравнением:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_P X_{t-P} + = Z_t. \qquad (1)$$

Или в более удобной форме записи:

$$X_t = Z_t - \sum_{n-1}^{P} a_n X_{t-n}, \qquad (2)$$

где $a_0, a_1, a_2, \ldots, a_P$ – постоянные коэффициенты, а Z_t – белый шум.

Несмотря на то, что система мониторинга предоставляет набор данных в полном объеме, значения приведенных в (1) коэффициентов неизвестны. А между тем, именно они определяют влияние каждого предиктора, используемого при расчете параметров модели. Другими словами, коэффициенты измеряют предельные эффекты переменных предиктора.

Для того, что выполнялось условие эргодичности, корни характеристического уравнения должны лежать вне единичного круга. Если это условие не выполняется, то значения членов ряда неограниченно возрастают.

Актуальной является задача идентификации модели согласно системе Юла-Уокера. Решение системы уравнений для оценки параметров фильтра и дисперсии сложная вычислительная задача. Для возможности ее разрешения в настоящей статье предложено использование программы математического моделирования с тем, чтобы не только получить оценочные характеристики, но и сравнить их с классическим подходом. Рассматриваемая задача оценивания решается также как и для случая многомерной регрессии. Схема оценивания параметров авторегрессии в классическом подходе подразумевает минимизацию критерия суммы квадратов ошибок. Необходимо отметить, что в практической реализации не так часто пользуются моделями с порядком больше двух. Для нахождения адекватного порядка авторегрессионной зависимости рассчитываются частные автокорреляционные функции, которые отражают степень линейной корреляции между наблюдениями X_t и X_{t-P}, при том, что влияние промежуточных членов ряда $X_{t-1}, X_{t-2}, \ldots, X_{t-P+1}$ на X_t устранено.

Для авторегрессионного процесса первого порядка все частные автокорреляции равны нулю. Таким образом, каждое наблюдение зависит только от предыдущего и соответственно, процессы называются Марковскими. Для рассматриваемого в статье авторегрессинного ряда порядка р частные корреляции, которые представляют собой корреляции между наблюдениями X_t и X_{t-P} при устранении влияния промежуточных членов ряда, будут равны последнему коэффициенту характеристического уравнения авторегрессионного процесса, а частные корреляции более высокого порядка равны нулю.

Теоретические исследования

Возможностью оценки параметров фильтра и дисперсии является расчет эмпирических уравнений Юла-Уокера. Для этого расчета требуется ковариационная функция авторегрессионного процесса, которую можно оценить по ряду наблюдений с помощью выборочной ковариационной функции. Выборочная ковариационная функция смещена, но существует также несмещенная оценка, называемая модифицированной выборочной ковариационной функцией [5-7]. Выборочную ковариационную функцию можно рассчитать в комплексной форме:

$$\hat{c}_{XX}(m) = \frac{1}{N} \sum_{i=1}^{N-|m|} x_i + |m| x_i^*, \text{ если } \hat{\mu}_X \neq 0 \qquad (3)$$

$$\hat{c}_{XX}(m) = \frac{1}{N} \sum_{i=1}^{N-|m|} (x_{i+-|m|} - \hat{\mu}_X)(x_i + \hat{\mu}_X)^*, \text{ если } \hat{\mu}_X \neq 0 \qquad (4)$$

где N – количество наблюдений; m – целочисленное значение на интервале, если $-(N-1) \leq m \leq (N-1)$; $\hat{\mu}_X$ – оценочное математическое ожидание последовательности $x = (x_1, x_2, \ldots, x_N)$.

Выборочная ковариационная функция является четной функцией, как и сама ковариационная функция. При N – количестве наблюдений мы получаем уже всю информацию об оценке в последовательность $\hat{c}_{XX}(0), \hat{c}_{XX}(1), \hat{c}_{XX}(N-1)$. Выборочные отклонения измеряются по отношению к выборочным средним значениям величин и имеют тенденцию к занижению отклонений от истинных средних значений.

Модифицированную выборочную ковариационную функцию (несмещенную) можно рассчитать согласно следующему выражению:

$$\hat{c}_{XX}(m) = \frac{1}{N-|m|}\sum_{i=1}^{N-|m|} x_i + |m| x_i^*, \text{ если } \hat{\mu}_X \neq 0 \quad (5)$$

$$\hat{c}_{XX}(m) = \frac{1}{N-|m|}\sum_{i=1}^{N-|m|} (x_{i+-|m|} - \hat{\mu}_X)(x_i + \hat{\mu}_X)^*, \text{ если } \hat{\mu}_X \neq 0 \quad (6)$$

Для целей регрессионного анализа пользуются определением выборочной дисперсии для выборки из ряда наблюдений.

При соответствующей оценке ковариационной функции эмпирические уравнения Юла-Уокера могут быть рассчитаны с соответствующими оценками [8-11]:

$$\hat{c}_{XX}(m) + \sum_{m=1}^{л} \hat{a}_n \hat{c}_{XX}(m-n) = \hat{c}_{ZX}(m) = \sigma_Z^2 \sigma_m = \begin{cases} \sigma_Z^2 & m=0 \\ 0 & m>0 \end{cases}.$$

При m>0 получаем систему уравнений:

$$\sum_{n=1}^{k} \hat{a}_n \hat{c}_{XX}(1-n) = -\hat{c}_{ZX}(1);$$

$$\sum_{n=1}^{k} \hat{a}_n \hat{c}_{XX}(2-n) = -\hat{c}_{ZX}(2);$$

$$\dots$$

$$\sum_{n=1}^{k} \hat{a}_n \hat{c}_{XX}(k-n) = -\hat{c}_{ZX}(k).$$

В матричной форме:

$$\begin{pmatrix} \hat{c}_{XX}(0) & \hat{c}_{XX}(1) & \cdots & \hat{c}_{XX}(k-1) \\ \hat{c}_{XX}(1) & \hat{c}_{XX}(0) & \cdots & \hat{c}_{XX}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_{XX}(k-1) & \hat{c}_{XX}(k-2) & \cdots & \hat{c}_{XX}(0) \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_k \end{pmatrix} = -\begin{pmatrix} \hat{c}_{XX}(1) \\ \hat{c}_{XX}(2) \\ \vdots \\ \hat{c}_{XX}(k) \end{pmatrix}$$

$$N \geq k+1. \quad (7)$$

Необходимо только k +1 – количество наблюдений, чтобы получить $\hat{c}_{XX}(0), \hat{c}_{XX}(1), \hat{c}_{XX}(k)$.

Дисперсию можно оценить, используя эмпирическое уравнение Юла-Уокера:

$$\hat{\sigma}_Z^2 = \hat{c}_{XX}(0) + \sum_{n=1}^{k} \hat{a}_n \hat{c}_{XX}(n). \quad (8)$$

Экспериментальные исследования

Экспериментальные исследования заключаются в программной реализации изложенных теоретических выводов и нахождении параметров авторегрессионного процесса и дисперсии через решение системы уравнений Юла-Уокера.

Первоначально, в программе математического моделирования была написана функция *covfct* для определения смещенной и несмещенной ковариационной функции выборки. Для расчета обеих выборочных ковариационных функций были взяты 200 случайных чисел. Кроме того, возможности программы математического моделирования позволяют получить ковариации, пользуясь встроенной функцией Matlab *xcov*. На рисунке 1 показаны результаты программного моделирования: смещенная ковариационная функция (верхний график) и модифицированная несмещенная ковариационная функция (нижний график).

Смещенная выборочная ковариационная функция нормирована на $1/N$ и постоянна со сдвигом m, согласно (3). В отличие от этого модифицированная выборочная ковариационная функция (несмещенная) нормализована на $\frac{1}{N-|m|}$ согласно (5).

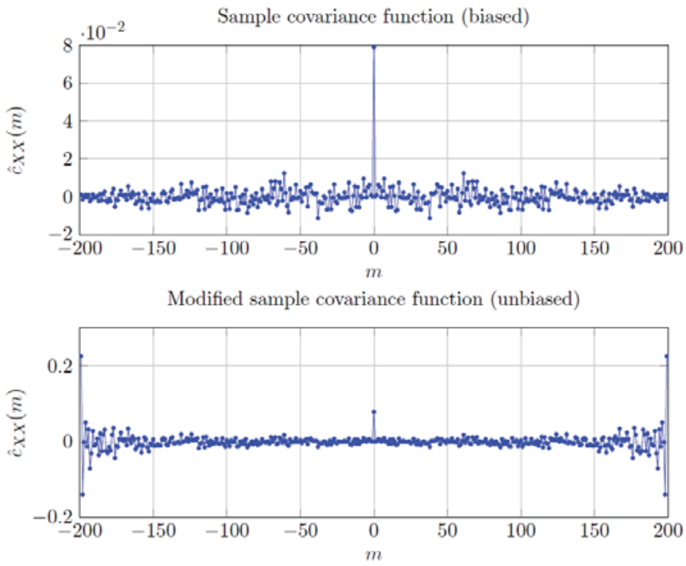

Рис. 1. Пример ковариационной функции

Несмещенная оценка достигается путем умножения выборочной оценки на $\frac{N}{N-|m|}$. Если m большое, то коэффициент $\frac{1}{N-|m|}$ уменьшается, а значения на границах увеличиваются (рис. 1).

Кроме того, выборочная ковариационная функция также является показателем, если последовательность коррелирована. Для последовательности дискретного белого шума максимальное значение находится при $\hat{c}_{XX}(0)$, поскольку каждое значение x_i практически не зависимы друг от друга. Таким образом, последовательность случайных чисел можно интерпретировать как реализацию дискретного белого шума.

Для генерации авторегрессионного процесса необходимо отфильтровать белый шум с помощью рекурсивного цифрового фильтра. Программная реализация АР-процесса осуществляется через функцию «фильтр» в программе математического моделирования. Порядок авторегрессионного процесса задается равным 5, и тогда, коэффициенты фильтра $a_0, \dots, a_5$. Таким образом, получается последовательность авторегрессионного процесса с определенными коэффициентами фильтра, которые и представляют собой задачу определения их значений, и последовательностью белого шума.

Теперь, для задачи идентификации модели авторегрессии требуется определить первые 11 значений выборочной ковариационной функции, $\hat{c}_{XX}(0), \hat{c}_{XX}(1), \dots, \hat{c}_{XX}(10)$. Это может быть выполнено в среде Matlab с использованием функции *covfct*. Затем проводится оценка параметров, коэффициентов фильтра, a_i и дисперсии $\hat{\sigma}_Z^2$ авторегрессионного процесса через эмпирическое уравнение Юла-Уокера с помощью алгоритма исключения Гаусса.

Для того, чтобы сгенерировать матрицу коэффициентов, используется функция toeplitz в программе математического моделирования.

Ковариационная функция может быть оценена с помощью выборочной ковариационной функции с использованием *covfct*. В таблице 1 приведены значения ковариационной функции авторегрессионного процесса.

Таблица 1

Расчетные значения ковариационной функции авторегрессионного процесса

Параметр	Значение
$\hat{c}_{XX}(0)$	2.1005
$\hat{c}_{XX}(1)$	-0.8592
$\hat{c}_{XX}(2)$	-0.2070
$\hat{c}_{XX}(3)$	0.7895
$\hat{c}_{XX}(4)$	-1.3956
$\hat{c}_{XX}(5)$	0.4238
$\hat{c}_{XX}(6)$	0.4821
$\hat{c}_{XX}(7)$	-0.6743
$\hat{c}_{XX}(8)$	0.8611
$\hat{c}_{XX}(9)$	-0.1827
$\hat{c}_{XX}(10)$	-0.4685

Система уравнений с эмпирическим уравнением Юла-Уокера (7) может быть рассчитана с использованием алгоритма исключения Гаусса с $O(k^3)$ плавающей точкой в соответствии с теоретическими выкладками.

В таблице 2 приведены оценочные коэффициенты AR(p) – процесса, решающего эмпирическое уравнение Юла-Уокера с помощью алгоритма исключения Гаусса.

Таблица 2

Оценочные коэффициенты AR (p) –процесса

Коэффициент	Расчетное значение	Ошибка
$\hat{a}_1$	0.4766	0.0234
$\hat{a}_2$	0.2881	0.0119
$\hat{a}_3$	0.0913	0.0087
$\hat{a}_4$	0.6695	0.0305
$\hat{a}_5$	0.2895	0.0105
	Σ	0.0850

Дисперсию также можно оценить, решив эмпирическое уравнение Юла-Уокера (8) $\hat{\sigma}_Z^2 = 0.8918$.

Полученные оценочные значения коэффициентов рекурсивного цифрового фильтра и дисперсии максимально приближены к их действительным значениям.

Заключение

Для проверки адекватности авторегрессионных моделей можно использовать программы математического моделирования, разработанные для моделей множественной регрессии на основании того, что задачи оценивания параметров множественной линейной регрессии по методу наименьших квадратов не имеют существенных различий с авторегрессией p-го порядка. При этом, частные корреляции между компонентами ряда, отстоящими друг от друга более чем на пять тактов, равны нулю.

Рассмотренная методология представляет интерес поскольку современные инфокоммуникационные системы представляют собой комплексные системы со множеством состояний и со-зависимостями между ними. Поэтому задача идентификации модели посредством программы математического моделирования упрощает процесс нахождения параметров авторегрессионного ряда и позволяет устранить недостатки классических методов.

В данной статье задача идентификации модели решалась через решение эмпирических уравнений Юла-Уокера с помощью алгоритма исключения Гаусса. Необходимо отметить, что по сравнению с классическим методом наименьших квадратов, алгоритм Гаусса позволяет работать с моделями более высоких порядков. Данный подход особенно результативен в случае долгосрочного прогнозирования. Механизмы поведения системы со временем меняются, вносятся сезонные коррективы, меняется тренд данных, поэтому и параметры моделей не остаются неизменными. Алгоритм, лежащий в основе программной реализации, позволяет уточнить известные подходы к прогнозированию, основанные на анализе временных рядов, и улучшить ключевые показатели качества функционирования системы.

За прошедшие годы признание прогнозирования как науки возросло, равно как и признание его необходимости. Что еще более важно, качество объективных прогнозов также улучшилось, что является прямым результатом признания того, что улучшения в оценочном прогнозировании могут быть достигнуты путем применения хорошо структурированных и систематических подходов.

Литература

1. *Шерстнева О.Г., Шерстнева А. А.* (2020). Анализ сети связи с учетом показателей надежности. *Вестник РГРТУ*, 73. С. 52-58.
2. *Шерстнева А.А.* (2020). Анализ и прогнозирование параметров авторегрессионного процесса p-го порядка. *Вестник ИжГТУ им. М. Т. Калашникова*. 23(4). С. 77-84.
3. *Домбровский В.В., Объедко Т.Ю.* (2019). Оптимальные стратегии прогнозирующего управления системами со случайными параметрами, описываемыми многомерной регрессионной моделью с марковским переключением режимов. *Вестн. Том. гос. ун-та*, УВТиИ, 48. С. 4-12.
4. *Домбровский В.В., Пашинская Т.Ю.* (2018). Прогнозирующее управление системами с марковскими скачками и авторегрессионным мультипликативным шумом с марковским переключением режимов. *Вестн. Том. гос. ун-та*. УВТиИ, 44. С. 4-9.
5. *H. Wickham.* (216). *Elegant graphics for data analysis.* 2td ed. Springer. 213 p. (in English)
6. *G. Athanasopoulos, R.J. Hyndman, N. Kourentzes, F. Petropoulos.* (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research, 262*(1), pp.60-74. (in English)
7. *C. Bergmeir, R.J. Hyndman, J.M. Benítez.* (2016). Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *International Journal of Forecasting, 32*(2), pp. 303-312. (in English)
8. *C. Bergmeir, R.J. Hyndman, B. Koo.* (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis, 120,* pp.70-83. (in English)
9. *S.L. Wickramasuriya, G. Athanasopoulos.* (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J American Statistical Association, 114*(526), pp. 804-819. (in English)
10. *F.E. Harrell.* (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis.* 2nd ed. New York, USA: Springer. 568 p. (in English)
11. *K. Madsen, H.B. Nielsen, O. Tingleff.* (2004). Methods for Non-linear Least Squares Problem Cobenhavn, Technical University of Denmark, 2004. 30 p.

IDENTIFICATION OF AN AUTOREGRESSIVE PROCESS USING MATLAB

Alina A. Sherstneva, *SibSUTIS, Novosibirsk, Russia,* **shers7neva@gmail.com**
Olga G. Sherstneva, *SibSUTIS, Novosibirsk, Russia,* **sherstneva@ngs.ru**

Abstract
The article aims to apply the methods of regression analysis in estimation task the reliability indicators of an infocommunication system and data trend forecasting. The task of assessment the reliability indicators of infocommunication systems is based, traditionally, on statistical data, the collection and processing is carried out by monitoring system. To obtain the calculated indicators as close as possible to the real practical results it is necessary to process a large number of measurements. In this sense, the theory of filtration is widely used in various estimation problems. It allows to support the possibility of effective solution of technical issues and implementation through mathematical modeling programs. The article is aimed to consider the issues of data trend forecasting for parameters calculation of infocommunication systems. One of the most effective solutions is the use of regression analysis methods. The research gives development and analysis of mathematical and program models. Based on theoretical calculations, methods for solving the problem are determined and experimental research are carried out. The article solves the problem of identifying an autoregressive process through the Yule-Walker equations. In addition to theoretical calculations, a program has been developed that allows handle the process of calculations in the Matlab environment. As a result, it is proposed to compare the results of identification time series of variables through the Yule-Walker system with the classical parameter estimation by the least squares method. The results are shown graphically.

Keywords: *regression analysis; least squares estimation; forecasting; data trend; identification; autoregressive process.*

References

1. O.G. Sherstneva, A.A. Sherstneva (2020). Analiz seti svyazi s uchetom pokazatelej nadezhnosti. *Vestnik RGRTU*, 73, pp. 52-58.
2. A.A. Sherstneva (2020). Analiz i prognozirovanie parametrov avtoregressionnogo processa r-go poryadka. *Vestnik IzhGTU im. M. T. Kalashnikova.* 23(4), pp. 77-84.
3. V.V. Dombrovskij, T.Yu. Ob"edko (2019). Optimal'nye strategii prognoziruyushchego upravleniya sistemami so sluchajnymi parametrami, opisyvaemymi mnogomernoj regressionnoj model'yu s markovskim pereklyucheniem rezhimov. *Vestn. Tom. gos. un-ta, UVTiI*, 48, pp. 4-12.
4. V.V. Dombrovskij, T.Yu. Pashinskaya (2018). Prognoziruyushchee upravlenie sistemami s markovskimi skachkami i avtoregressionnym mul'tiplikativnym shumom s markovskim pereklyucheniem rezhimov. *Vestn. Tom. gos. un-ta. UVTiI*, 44, pp. 4-9.
5. H. Wickham. (2016). Elegant graphics for data analysis. 2td ed. Springer. 213 p.
6. G. Athanasopoulos, R.J. Hyndman, N. Kourentzes, F. Petropoulos. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research,* 262(1), pp. 60-74.
7. C. Bergmeir, R.J. Hyndman, J.M. Benetez. (2016). Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. *International Journal of Forecasting,* 32(2), pp. 303-312.
8. C. Bergmeir, R.J. Hyndman, B. Koo. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis,* 120, pp. 70-83. (in English)
9. S.L. Wickramasuriya, G. Athanasopoulos. (2019). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J American Statistical Association*, 114(526), pp. 804-819.
10. F.E. Harrell. (2015). Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. New York, USA: Springer. 568 p.
11. K. Madsen, H.B. Nielsen, O. Tingleff. (2004). Methods for Non-linear Least Squares Problem Cobenhavn, Technical University of Denmark, 2004. 30 p.

Information about authors:
Alina A. Sherstneva, *Candidate of Tech. Sciences, associated professor, Siberian State University of Telecommunications and Information Sciences, Department of Electrical Communication, Novosibirsk, Russia*
Olga G. Sherstneva, *SibSUTIS, Novosibirsk, Russia, associated professor, Siberian State University of Telecommunications and Information Sciences, Department of Electrical Communication, Novosibirsk, Russia*