

**ПРОГРАММА ОПТИМИЗАЦИИ ПАРАМЕТРОВ РЕГРЕССИОННЫХ
МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ЛЕВЕНБЕРГА – МАРКВАРДТА.**
Рыбаков В.Н.

Институт биоорганической химии НАН Беларуси

Анализ кинетических данных в фармакологии требует программного обеспечения. В данном случае, для определения параметров уравнений заданных в явном виде написана программа на VBA в виде макроса согласно следующему итерационному процессу

$$K_{k+1} = K_k + s(F^T F + dI)^{-1}(F^T(Y - Y^t)),$$

где, K – вектор параметров, Y – вектор экспериментальных данных, Y^t – вектор вычисленных значений согласно уравнению приближения, F – матрица $m \times n$ первых частных производных функции по параметрам, F^T – транспонированная матрица, I – единичная матрица, k – номер итерации, m – число данных, n – число параметров, s – шаг итерации, d – поправка Левенберга – Марквардта [1,2,3]. По определению $\sigma^2 = SS/(m-n)$, где SS – сумма квадратов отклонений. Если $A = (F^T F)^{-1}$, где A – обратная матрица $n \times n$, а D – вектор элементов главной диагонали A , то вектор стандартных ошибок параметров $SD = \sigma(Abs(D))^{1/2}$. Первые частные производные, как правило, вычисляются в аналитическом виде или, если необходимо, аппроксимируются методом центральных разностей

$$DE = (Y^t_{(K-2q)} - 8Y^t_{(K-q)} + 8Y^t_{(K+q)} - Y^t_{(K+2q)})/12/q,$$

где, DE – вектор частных производных, $q = K/10000$. Если $s \geq Abs(SS_k - SS_{(k+1)})/SS_k$ итерации заканчиваются. Критерий останова s , обычно, устанавливается равным нулю. Тестирование макроса выполнено на нелинейных уравнениях различной сложности, которые представлены на сайте NIST, (Национальный Институт Стандартов и Технологий), США [4]. Здесь, уравнение, число данных, предикторы, два вектора начальных значений параметров составляют одну из 27 тестовых задач. Все сертифицированные значения параметров, стандартных ошибок параметров, сумм квадратов отклонений и сигм (σ) даны в стандартном виде с 11 значащими цифрами. Результаты тестирования макроса приведены в табл. 1.

Таблица 1. Число верных значащих цифр в величинах параметров, стандартных ошибок и SS , вычисленных с применением макроса, с двумя начальными векторами параметров.

Модель	Слож - ность модели	Число пара - метров	Число верных значащих цифр							
			Старт 1				Старт 2			
			параметр	SD	SS	k	параметр	SD	SS	k
Misra1a	низкая	2	11	11	11	42	11	11	11	23
Chwirut2	низкая	3	11	11	11	9	11	11	11	10
Chwirut1	низкая	3	11	11	11	15	11	11	11	15
Lanczos3	низкая	6	11	9	11	474	11	9	11	386
Gauss1	низкая	8	11	11	11	20	11	11	11	12
Gauss2	низкая	8	11	11	11	20	11	11	11	11
DanWood	низкая	2	11	11	11	22	11	11	11	20
Misra1b	низкая	2	11	11	11	27	11	11	11	32
Kirby2	средняя	5	11	11	11	27	11	11	11	29
Hahn1	средняя	7	11	11	11	67	11	11	11	35
Nelson	средняя	3	11	11	11	132	11	11	11	30
MGH17	средняя	5	11	10	11	99	11	10	11	37
Lanczos1	средняя	6	11	2	2	438	11	2	2	453
Lanczos2	средняя	6	11	9	10	197	11	9	10	127
Gauss3	средняя	8	11	11	11	27	11	11	11	21
Misra1c	средняя	2	11	11	11	42	11	11	11	29
Misra1d	средняя	2	11	11	11	28	11	11	11	22

Roszman1	средняя	4	11	11	11	42	11	11	11	41
ENSO	средняя	9	11	11	11	47	11	11	11	40
MGH09	высокая	4	11	11	11	95	11	11	11	29
Thurber	высокая	7	11	11	11	45	11	11	11	25
BoxBOD	высокая	2	11	11	11	18	11	11	11	18
Rat42	высокая	3	11	11	11	24	11	11	11	24
MGH10	высокая	3	11	10	11	7738	11	11	11	207
Eckerle4	высокая	3	11	11	11	33	11	11	11	19
Rat43	высокая	4	11	11	11	41	11	11	11	25
Bennett5	высокая	3	11	7	11	522	11	7	11	1063

Примечание. k – число итераций, старт 1 и старт 2 – два различных вектора начальных значений параметров.

В табл. 2 представлены сертифицированные и полученные с помощью макроса значения параметров, стандартных ошибок и SS для уравнения MGH10 с выбором проблемного начального вектора параметров, старт 1. Считается, что это задача является трудной даже для очень хороших алгоритмов.

Таблица 2. Тестирование макроса, уравнение MGH10.

Старт 1	Сертифицированные значения		Полученные значения	
	параметр	SD	параметр	SD
2	5.6096364710E-03	1.5687892471E-04	5.6096364710E-03	1.5687892470E-04
400000	6.1813463463E+03	2.3309021107E+01	6.1813463463E+03	2.3309021106E+01
25000	3.4522363462E+02	7.8486103508E-01	3.4522363462E+02	7.8486103505E-01
	SS = 8.7945855171E+01		SS = 8.7945855171E+1	

В итоге, полученные значения отличаются от сертифицированных лишь в последнем знаке стандартных ошибок. Результаты тестирования макроса не отличаются от данных приведенных в работе [5].

Сравнение макроса с известной программой GraphPad Prism 6.02 сделано на основе эксперимента по ингибированию общей АТФ-азы сарколеммы сердца препаратом окисленного крахмала, который любезно предоставлен Дарашкевичем О.Н. Сарколемма выделена из гомогената левых желудочков крыс в градиенте плотности сахарозы на центрифуге L8-50 M/E в роторе 50 TI. Активность общей АТФ – азы определена в среде: Tris·HCl – 50 mM, NaCl – 88 mM, KCl – 10 mM, MgCl₂ – 3 mM, MgATP – 3 mM, pH 7.5, 37⁰C. Реакция инициировалась MgATP и заканчивалась добавлением SDS (конечная концентрация – 0.33%). Неорганический фосфат определен методом Taussky и поглощение проб измерено на спектрофотометре SPECORD M – 40 при 737 нм. Активность фермента выражается в ΔAU между опытом и контролем за 10 мин хода реакции. Зависимость активности общей АТФ – азы сарколеммы от концентрации препарата аппроксимирована уравнением Хилла

$$Y^t = K1 - K1[X]^{K3}/(K2^{K3} + [X]^{K3}), \quad (1)$$

где, Y^t – активность фермента, K1 – активность фермента в отсутствии препарата, [X] – относительная концентрация препарата, K2 – относительная концентрация препарата, при которой Y^t = K1 / 2, K3 – коэффициент Хилла. Данные показаны в табл. 3 и на рис. 1.

Таблица 3. Значения GraphPad Prism 6.02 и макроса.

Начальный вектор	GraphPad Prism 6.02		Макрос	
	параметр	SD	параметр	SD
K1 = 1	0.499945	0.024815	5.0112010001E-01	2.4858542531E-02
K2 = 1	2.798558	0.172946	2.7925348756E+00	1.7144965650E-01
K3 = 1	5.325816	0.826724	5.3006611610E+00	8.1176817913E-01
	SS = 0.004665		SS = 4.6543132949E-03	

Незначительное расхождение в цифрах имеет место вследствие того, что сумма квадратов отклонений у макроса меньше в сравнении с таковой у GraphPad Prism 6.02. Следует заметить, что первые частные производные для этой модели вычисляются методом центральных разностей, а SS минимизируется с учетом постоянной относительной ошибки, т.е. находится взвешенная сумма квадратов $SS_w = \sum w(Y - Y^t)^2$, где $w = 1/ Y^t / Y^t$. Критерий останова $s = 0$. Число итераций = 37.

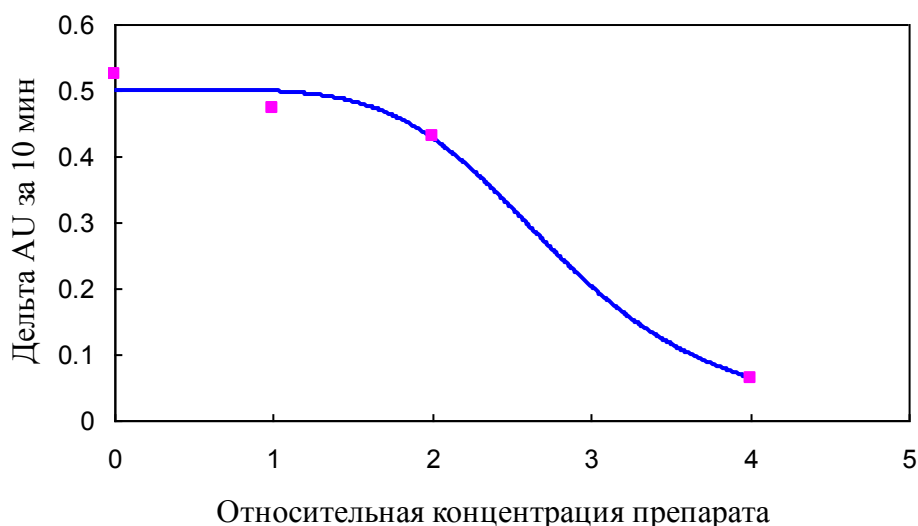


Рис. 1. Зависимость активности общей АТФ-азы сарколеммы от относительной концентрации препарата. ■ – экспериментальные точки нанесены на теоретическую кривую, сплошную линию, вычисленную согласно уравнению (1).

Данные по тестирование макроса показали очень хорошее соответствие со значениями задач NIST, имеют почти полное совпадение с результатами [5] и программой GraphPad Prism 6.02.

Литература

1. Демиденко Е.З. Оптимизация и регрессия. М. 1989.
2. Эберт К., Эдерер Х. Компьютеры. Применение в химии. М. 1988.
3. Madsen K., Nielsen H.B., Tingleff O. Methods for non – linear squares problems. Technical University of Denmark. 2004.
4. www.itl.nist.gov/div898/strd/nls/nls_main.shtml. NIST StRD Nonlinear Least Squares Regression Datasets.
5. rsc.chemometrics.ru/papers/thesis_alp.pdf. Померанцев А.Л. Методы нелинейного регрессионного анализа для моделирования кинетики химических и физических процессов. М. 2003.

OPTIMIZATION PARAMETERS PROGRAM OF REGRESSION MODEL USING THE LEVENBERG – MARQUARDT ALGORITHM.

Ribakov V.N.

The Institute of bioorganic chemistry NAS of Belarus.

The paper describes a program that solves the problem of optimizing the parameters of the regression models by least squares method using the Levenberg – Marquardt algorithm. The program is written in VBA as a macro, testing is carried out on the 27 tasks NIST. The test results showed very good agreement with the data problems NIST.