

Mieczysław F. PAZDUR, Anna PAZDUR

ZASTOSOWANIE NIELINIOWEJ METODY NAJMNIEJSZYCH KWADRATÓW
DO ESTYMACJI PARAMETRÓW FUNKCJI APROKSIMUJĄCEJ WYNIKI POMIARÓW

Streszczenie. W artykule przedyskutowano zagadnienie wygładzenia wyników pomiarów w przypadku, gdy nieznanne parametry funkcji wygładzającej występują w postaci nieliniowej. Przedstawiono modyfikację jednej z metod wyznaczania estymatorów (tzw. metody najszybszego spadku), polegającą na linearyzacji warunku minimalizacji sumy kwadratów odchyłeń punktów doświadczalnych od funkcji wygładzającej. Podano skrócony opis algorytmu obliczeń dla wyznaczania estymatorów parametrów funkcji wygładzającej będącej sumą kilku eksponent.

1. Wstęp

Rozwój techniki obliczeniowej, związany z pojawieniem się szybkich maszyn cyfrowych, wyposażeń w obszerne biblioteki programów, znacznie rozszerzył klasę zagadnień związanych z opracowaniem wyników eksperymentów. W szczególności wygładzanie wyników pomiarów przy użyciu wielomianów stało się zagadnieniem raczej trywialnym ze względu na rozpowszechnienie programów bibliotecznych realizujących interpolację wielomianową. Z fizycznego punktu widzenia wygładzanie wyników pomiarów przy pomocy wielomianów tylko w nielicznych przypadkach ma głębsze znaczenie i w zasadzie powinno być traktowane jako pierwszy krok w kierunku opisanie badanego procesu i wyznaczenia charakterystycznych go parametrów. Większość procesów fizycznych opisywana jest równaniami, które prowadzą do bardziej skomplikowanych zależności niż zależność wielomianowa. Dla spełnienia naturalnego żądania - aby funkcja użyta do opisu wyników pomiarów miała jasną interpretację fizyczną, zgodną z przewidywaniami teoretycznymi - konieczne jest rozważanie funkcji, w których zarówno zmienne, jak i parametry mogą występować w postaci nieliniowej.

Przy pomiarach związanych z zastosowaniami krótkożyłowych izotopów promieniotwórczych proces rozpadu promieniotwórczego prowadzi w naturalny sposób do funkcji aproksymującej o postaci

$$y = B + Ae^{-\lambda t} \quad (1)$$

gdzie: parametry A i B występują liniowo, a λ nieliniowo.

Aproksymacja wyników pomiarów funkcją postaci (1) ma podstawowe znaczenie przy wyznaczaniu stałych zaniku krótkożyciowych pierwiastków promieniotwórczych. Podobna matematycznie postać funkcji występuje przy badaniu procesów typu relaksacyjnego - wtedy $B = -A$ i funkcja (1) wyraża się wzorem

$$y = A(1 - e^{-\lambda t}). \quad (2)$$

W przypadku występowania kilku genetycznie związanych radioizotopów (będących członkami szeregu promieniotwórczego) funkcja (1) zawiera kilka eksponent

$$y = B + A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + \dots + A_n e^{-\lambda_n t} \quad (3)$$

gdzie współczynniki A_i , $i = 1, 2, \dots, n$ zależą od stałych zaniku λ_i oraz początkowych ilości atomów poszczególnych izotopów promieniotwórczych i w niektórych przypadkach mogą być od siebie zależne [1].

Identyczna postać funkcji aproksymującej występuje w zagadnieniach analizy aktywacyjnej - współczynniki A_i są w tym przypadku z reguły niezależne od siebie.

Robinson [2] i Sterlinski [3] rozpatrywali problem wyznaczenia stałej λ występującej we wzorze (1) metodą dopasowania prostej najmniejszych kwadratów do przekształconych wyników pomiarów otrzymanych po odjęciu znanego tła B , stwierdzając, że dopasowanie tego typu wprowadza systematyczny błąd, którego wielkość zależy od liczby punktów pomiarowych, liczby okresów połowicznego zaniku objętych pomiarami oraz od szybkości zliczeń tła. Rakowić i Prouza [4, 5] opisali metodę kolejnych przybliżeń dla numerycznego wyznaczenia parametrów funkcji (3) dla $n = 3$ (trzy niezależnie od siebie pierwiastki promieniotwórcze), wychodząc z założenia, że po dostatecznie długim okresie czasu rejestrowane promieniowanie pochodzi tylko od izotopu o najdłuższym okresie połowicznego zaniku.

Higbie [6] opisał algorytm wykorzystujący linearyzację funkcji $\exp(-\lambda t)$ względem niewielkich poprawek do znanych w przybliżeniu początkowych wartości λ_0 . Korzystając z wyznaczonych metodą graficzną wartości współczynników przybliżenia zerowego, otrzymuje się następnie, na drodze iteracyjnej, parametry "najlepszego dopasowania".

W obecnej pracy opisano modyfikację jednej z wersji otrzymywania estymatorów nieliniowych parametrów funkcji wykładającej (metody najszybszego spadku) oraz jej zastosowanie do wyznaczania estymatorów parametrów funkcji będącej sumą kilku eksponent.

2. Ogólne sformułowanie problemu wyznaczania estymatorów metoda najmniejszych kwadratów

Przyjmijmy, że w wyniku pomiarów otrzymano n par liczb (y_i, t_i) , $i = 1, \dots, n$, a zależność między tymi zmiennymi aproksymujemy funkcję

$$y = f(\hat{a}; t) \quad (4)$$

gdzie: $\hat{a} = (a_1, \dots, a_m)$ jest m -wymiarowym wektorem parametrów ($m < n$), a t jest zmienną niezależną, o której zakłada się, że jej wartość dla każdego pomiaru jest znana dokładnie.

Powszechnie przyjętym kryterium znajdowania estymatorów parametrów \hat{a} jest minimalizacja sumy kwadratów odchyleń punktów doświadczalnych od wartości przewidywanych funkcją (4), czyli minimalizacja wyrażenia

$$Q = Q(\hat{a}) \stackrel{df}{=} \sum_{i=1}^n w_i (y_i - f(\hat{a}; t_i))^2 \quad (5)$$

gdzie: współczynniki w_i oznaczają wagi statystyczne poszczególnych pomiarów.

Warunek minimalizacyjny można zapisać w postaci układu m równań

$$\frac{\partial Q}{\partial a_j} = 0, \quad j = 1, 2, \dots, m, \quad (6)$$

który po uwzględnieniu (5) przyjmuje postać

$$\sum_i w_i (y_i - f_i) \frac{\partial f_i}{\partial a_j} = 0 \quad (7)$$

lub

$$\sum_i w_i y_i \frac{\partial f_i}{\partial a_j} = \sum_i w_i f_i \frac{\partial f_i}{\partial a_j} \quad (7a)$$

Jeżeli parametry a_j występują liniowo w funkcji $f(\hat{a}; t)$, wówczas układ równań (7) jest układem m równań liniowych; problem sprowadza się do klasycznej metody najmniejszych kwadratów, dyskutowanej w wielu podręcznikach i monografiach (por. np. [7-9]). Gdy parametry a_j występują w postaci nieliniowej, układ równań (7) jest również układem równań nieliniowych ze względu na szukane parametry a_j i tylko w nielicznych przypadkach może być rozwiązany w sposób ścisły.

W przypadku nieliniowym stosowane są następujące metody uzyskiwania estymatorów:

- a) metoda zamiany zmiennych [10],
- b) metoda linearyzacji funkcji aproksymującej względem nieliniowych parametrów (tzw. linearyzacja modelu),
- c) metoda najszybszego spadku (linearyzacja sumy kwadratów odchyłeń Q),
- d) metoda Monte Carlo.

Metoda zamiany zmiennych polega na znalezieniu takiego zespołu nowych zmiennych i parametrów, aby zagadnienie sprowadzić do przypadku liniowej metody najmniejszych kwadratów. Klasa zagadnień praktycznych, jakie mogą być w ten sposób rozwiązywane, jest jednak stosunkowo ograniczona.

Metody linearyzacyjne b) i c) mają charakter procedur iteracyjnych, w których zadaje się pewne przybliżone wartości początkowe szukanych parametrów i w kolejnych krokach obliczeń uzyskuje się wartości coraz dokładniejsze.

Metoda linearyzacji modelu w pewnych przypadkach może nie być zbieżna, metoda najszybszego spadku jest teoretycznie zawsze zbieżna, jednak niekiedy zbieżność ta może być bardzo powolna. Najlepsze cechy obu metod linearyzacyjnych łączy w sobie technika zwana kompromisem Marquardta [11, 12].

3. Modyfikacja metody najszybszego spadku

W ujęciu klasycznym metoda najszybszego spadku polega na linearyzacji sumy kwadratów odchyłeń, czyli funkcji Q , ze względu na parametry a_j . Zakłada się więc, że

$$Q = Q_0 + \sum_{j=1}^m \left(\frac{\partial Q}{\partial a_j} \right)_0 (a_j - a_j^0) + \dots \quad (8)$$

gdzie: a_j oznacza wartości parametrów a_j w zerowym przybliżeniu, $Q_0 = Q(a_1^0, a_2^0, \dots, a_m^0)$ oraz $\left(\frac{\partial Q}{\partial a_j} \right)_0$ oznacza wartość pochodnej otrzymaną po wstawieniu parametrów przybliżenia zerowego.

Opisane niżej odmienne metody najszybszego spadku polega na linearyzacji warunku minimalizacji sumy kwadratów odchyłeń punktów pomiarowych od wartości przewidywanych przez dopasowywaną funkcję

$$\frac{\partial Q}{\partial a_j} = \left(\frac{\partial Q}{\partial a_j} \right)_0 + \sum_{k=1}^m \left(\frac{\partial^2 Q}{\partial a_j \partial a_k} \right)_0 (a_k - a_k^0) = 0 \quad (9)$$

Korzystając z definicji funkcji Q (równ. (5)), otrzymuje się, po wstawieniu do (9) i uporządkowaniu, układ równań, który w zapisie macierzowym przyjmuje szczególnie prostą postać [13]

$$\hat{M} \Delta \hat{a} = \hat{V} \quad (10)$$

gdzie: \hat{M} oznacza macierz kwadratową $m \times m$, której elementy określone są wyrażeniami

$$M_{jk} = \sum_{i=1}^n \left\{ \left(\frac{\partial f_i}{\partial a_j} \right)_0 \left(\frac{\partial f_i}{\partial a_k} \right)_0 - (y_i - f_i) \left(\frac{\partial^2 Q}{\partial a_j \partial a_k} \right)_0 \right\} w_i, \quad (11)$$

\hat{V} jest m -wymiarowym wektorem o elementach

$$V_j = \sum_{i=1}^n (y_i - f_i) \left(\frac{\partial f_i}{\partial a_j} \right)_0 w_i \quad (12)$$

a $\Delta \hat{a}$ jest wektorem poprawek do parametrów a_j

$$\Delta a_j = a_j - a_j^0 \quad (13)$$

Rozwiązanie równania (10) ze względu na nowe wartości parametrów a_j ma postać

$$\hat{a} = \hat{a}_0 + \hat{M}^{-1} \hat{V} \quad (14)$$

gdzie: \hat{M}^{-1} oznacza macierz odwrotną do \hat{M} .

Wartości parametrów a_j otrzymane ze wzoru (14) są następnie traktowane jako nowe wartości wyjściowe dla kolejnego kroku obliczeń. Proces iteracyjny przerywa się, jeżeli przyrosty parametrów Δa_j stają się bardzo małe, w praktyce jako kryterium przerwania iteracji używa się warunku

$$\max_j \frac{|a_j - a_j^0|}{|a_j^0|} < \varepsilon \quad (15a)$$

lub

$$\sum_{j=1}^m \frac{|a_j - a_j^0|}{|a_j^0|} < \varepsilon \quad (15b)$$

gdzie: ε jest ustaloną małą liczbą (rzędu 10^{-4} - 10^{-8}).

Oznaczając przez Q_{\min} znaną minimalną wartość sumy kwadratów odchyleń, można w analogii do liniowej metody najmniejszych kwadratów otrzymać wyrażenia na elementy s_{jk} macierzy kowariancji \hat{S} parametrów a_j

$$s_{jk}^2 = Q_{\min} M_{jk}^{-1} / [W(n - m)] \quad (16)$$

gdzie: $W = \sum w_i$ (gdy wszystkim pomiarom przypisuje się identyczne wagi statystyczne, czyli $w_i = 1$, wówczas $W = n$).

Elementy diagonalne tej macierzy równe są kwadratowi dyspersji s_j parametrów a_j , czyli

$$s_j = \sqrt{Q_{\min} M_{jj}^{-1} / [W(n - m)]}. \quad (17)$$

Korzystając z definicji współczynników korelacji, otrzymuje się na podstawie (16) i (17) wyrażenia na elementy r_{jk} macierzy korelacji \hat{R} dla parametrów a

$$r_{jk}^2 = \frac{|M_{jk}^{-1}|}{(M_{jj}^{-1} M_{kk}^{-1})^{1/2}} \quad (18)$$

Przedziały ufności dla estymatorów parametrów a_j oraz dla przewidywanych wartości funkcji aproksymującej można otrzymać, korzystając z faktu, że wielkość Q może być traktowana jak zmienna losowa, mająca rozkład χ^2 o $f = n - m$ stopniach swobody (por. [14]).

4. Skrócony opis algorytmu obliczeń dla estymacji parametrów funkcji będącej sumą eksponent

Opisaną w poprzednim punkcie zmodyfikowaną wersję metody najszybszego spadku zastosowano do znalezienia estymatorów parametrów funkcji będącej sumą kilku eksponent oraz funkcji typu relaksacyjnego. Dla zwięzłości zapisu oznaczmy przez N liczbę eksponent występujących w wyrażeniu na funkcję aproksymującą wyniki pomiarów i przyjmijmy, że funkcja ta wyraża się wzorem

$$y = f(a; t) = a_1 e^{-a_2 t} + \dots + a_{2l-1} e^{-a_{2l} t} + \dots + a_{2N-1} e^{-a_{2N} t}. \quad (19)$$

Wyrażenia na pierwsze pochodne cząstkowe można zapisać w zwartej postaci

$$\frac{\partial f}{\partial a_{2j-1}} = \exp(-a_{2j}t) \quad (20a)$$

$$\frac{\partial f}{\partial a_{2j}} = -a_{2j-1}t \exp(-a_{2j}t) \quad (20b)$$

dla $j = 1, 2, \dots, N$. Większość spośród drugich pochodnych cząstkowych równa się zeru, różne od zera są tylko

$$\frac{\partial^2 f}{\partial a_{2j}^2} = a_{2j-1}t^2 \exp(-a_{2j}t) \quad (21a)$$

$$\frac{\partial^2 f}{\partial a_{2j} \partial a_{2j-1}} = -t \exp(-a_{2j}t) \quad (21b)$$

dla $j = 1, 2, \dots, N$. Elementy macierzy \hat{V} i \hat{M} wyrażają się wzorami

$$v_{2j-1} = \sum w_i (y_i - f_i) \exp(-a_{2j}t_i) \quad (22a)$$

$$v_{2j} = \sum w_i (y_i - f_i) t_i \exp(-a_{2j}t_i) \quad (22b)$$

$$M_{2j, 2k} = \begin{cases} a_{2j-1} \sum w_i t_i^2 (y_i - f_i + a_{2j-1} \exp(-a_{2j}t_i)) \exp(-a_{2j}t_i) & \text{dla } k=j \\ a_{2j-1} a_{2k-1} \sum w_i t_i^2 \exp(-(a_{2j} + a_{2k})t_i) & \text{dla } k \neq j \end{cases} \quad (23a)$$

$$M_{2j-1, 2k-1} = \sum w_i \exp(-(a_{2j} + a_{2k})t_i) \quad (23b)$$

$$M_{2j, 2k-1} = \begin{cases} \sum w_i t_i \exp(-a_{2j}t_i) (y_i - f_i - a_{2j-1} t_i \exp(-a_{2j}t_i)) & \text{dla } k=j \\ -a_{2j-1} \sum w_i t_i \exp(-(a_{2j} + a_{2k})t_i) & \text{dla } k \neq j \end{cases} \quad (23c)$$

gdzie: $j, k = 1, 2, \dots, N$.

Po obliczeniu elementów macierzy \hat{M} i \hat{V} obliczane są elementy macierzy odwrotnej \hat{M}^{-1} , a następnie z równania (14) wartości poprawionych parametrów a_j , które są używane jako parametry przybliżenia zerowego w następnym kroku iteracji.

Szczegółowy opis programu opartego na podstawowym algorytmie oraz dyskusja jego działania i przykłady wyników będą omówione w oddzielnej pracy.

LITERATURA

- [1] Strzałkowski A., Wstęp do fizyki jądra atomowego, PWN, Warszawa (1969).
- [2] Robinson D.C., Nucl. Instruments Methods 79 (1970) 65.
- [3] Sterlinski S., Nucl. Instruments Methods 47 (1967) 329.
- [4] Rakovic M., Prouza Z., Atompraxis 13 (1967) 361.
- [5] Rakovic M., Prouza Z., Istopenpraxis 11 (1975) 413.
- [6] Higbie J., Nucl. Instruments Methods 105 (1972) 279.
- [7] Ralston A., Wstęp do analizy numerycznej, PWN, Warszawa (1974).
- [8] Brandt S., Metody statystyczne i obliczeniowe analizy danych, PWN, Warszawa (1974).
- [9] Draper N.R., Smith H., Analiza regresji stosowana, PWN, Warszawa (1973).
- [10] Box G.E.P., Tidwell P.W., Technometrics 4 (1962) 531.
- [11] Marquardt D.W., J.Soc. Ind. Appl. Math. 2 (1953) 431.
- [12] Marquardt D.W., Chem. Eng. Progr. 55, nr 6 (1959) 65.
- [13] White J.J., Appl. Phys. 5 (1974) 57.
- [14] Rogers D.W.O., Nucl. Instruments Methods 127 (1975) 253.

ПРИМЕНЕНИЕ НЕЛИНЕЙНОГО МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ ОЦЕНКИ ПАРАМЕТРОВ ФУНКЦИИ АПРОКСИМИРУЮЩЕЙ РЕЗУЛЬТАТЫ ИЗМЕРЕНИЙ

Резюме

В статье рассматривается вопрос сглаживания результатов измерений в том случае, когда неизвестные параметры сглаживающей функции выступают в нелинейной форме. Представлена модификация одного из методов установления оценок так называемый метод быстрого падения, состоящий в линеаризации условия минимализации суммы квадратов отклонений опытных точек от сглаживающей функции.

Представлено также краткое описание алгоритмов расчётов для установления оценок параметров сглаживающей функции, являющейся суммой нескольких экспонент.

APPLICATION OF NONLINEAR LEAST SQUARES METHOD TO THE ESTIMATION
OF UNKNOWN PARAMETERS OF SMOOTHING FUNCTION

S u m m e r y

The problem of smoothing of results of measurements is discussed for the case when unknown parameters of smoothing function appear in a non-linear form. A modification of steepest descent method is described, consisting of linearization of minimalization condition. Short description of the algorithm for estimation of parameters of function being the sum of exponential functions is also given.