ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1980

Seria: AUTOMATYKA z. 53 Nr kol. 649

Ewa BIELIŃSKA

UWAGI O ZBIEŻNOŚCI REKURENCYJNEJ METODY NAJMNIEJSZYCH KWADRATÓW (RMNK)

Streszczenie. W artykule omówiono właściwości eksploatacyjne algorytmu rekurencyjnej metody najmniejszych kwadratów. Rozważono możliwości przyspieszenia zbieżności algorytmu.

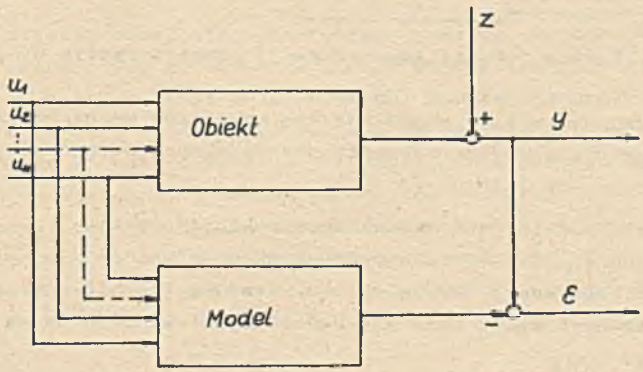
1. Wstęp

Mimo istnienia znacznej ilości metod pozwalających na bieżącą estymację parametrów modeli obiektów zakłócanych różnego rodzaju szumami, metoda najmniejszych kwadratów jest na razie najbardziej popularna. Identyfikowane tą metodą modele obiektów przedstawiają zależność zmiennej wyjściowej z obiektu od pomiarowo dostępnych zmiennych wejściowych, przy czym zależność ta musi być liniowa względem estymowanych parametrów. Metoda najmniejszych kwadratów stosowana jest do identyfikacji zarówno statycznych jak i dynamicznych modeli obiektów.

W przypadku identyfikacji modeli statycznych (rys. 1) równanie modelu przedstawione jest jako:

$$y = \underline{b}^T \underline{u} \tag{1.1}$$

Błąd identyfikacji ε powinien posiadać statystyczne właściwości takie jak zakłócenie działające na badany obiekt. Jest to możliwe w przypadku, gdy struktury równań obiektu i modelu są identyczne. Błąd identyfikacji interpretowany jest niejednokrotnie jako zakłócenie obiektu.



Rys. 1. Ogólny schemat identyfikacji charakterystyk statycznych

W przypadku identyfikacji modeli dynamicznych model obiektu zapisywany jest w postaci liniowego równania różnicowego:

$$y(k) = \sum_{i=1}^{n} a_i\, y(k-i) + \sum_{i=0}^{n} b_i\, u(k-p-i) \tag{1.2}$$

lub częściej za pomocą transmitancji "z"

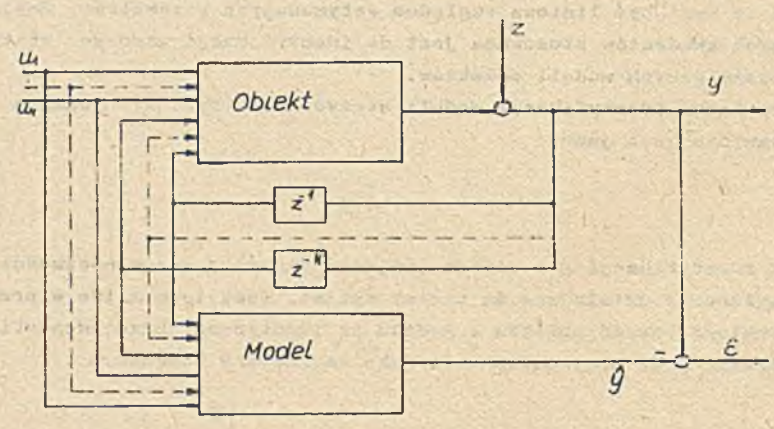
$$y(k) = z^{-p}\,\frac{B(z^{-1})}{A(z^{-1})}\,u(k) \tag{1.3}$$

gdzie:

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-n}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_n z^{-n}$$

p - jest znanym opóźnieniem obiektu.



Rys. 2. Ogólny schemat identyfikacji charakterystyk dynamicznych

Wykorzystując metodę najmniejszych kwadratów można uzyskać nieobciążone i najefektywniejsze estymaty poszukiwanych parametrów, przy spełnieniu następujących założeń [1],[2]:

1) struktura modelu jest zgodna ze strukturą obiektu,
2) błędy identyfikacji dla poszczególnych doświadczeń są statystycznie niezależne, posiadają zerową wartość średnią i stałą wariancję,
3) błędy identyfikacji oraz wielkości wejściowe obiektu są nieskorelowane ze sobą.

Przy założeniu, że zakłócenia działające na obiekt są stacjonarnymi procesami stochastycznymi, mogą one być przedstawione jako wynik liniowej filtracji białego szumu

$$z(k) = C(z^{-1})e(k) \tag{1.4}$$

gdzie: $e(k)$ jest białym szumem.
Dla spełnienia założenia 2 musi zachodzić:

$C(z^{-1}) = 1$, czyli zakłócenia obiektu mają charakter białego szumu.
Jeżeli - jak to ma miejsce w wielu praktycznych przypadkach - właściwości statystyczne zakłóceń odbiegają od właściwości białego szumu, stosowanie MNK daje obciążne estymaty parametrów modelu.

2. Rekurencyjna metoda najmniejszych kwadratów

Rekurencyjna metoda najmniejszych kwadratów pozwala na sekwencyjne wyliczanie wektora estymowanych parametrów w takt uzyskiwania nowych danych pomiarowych. Estymowany wektor parametrów $\hat{\underline{b}}_k$ wyliczany jest jako suma poprzednio uzyskanej estymaty $\hat{\underline{b}}_{k-1}$ uzupełnionej poprawką, wynikającą z błędu wyjścia jaki powstałby przy zastosowaniu w momencie k estymaty $\hat{\underline{b}}_{k-1}$

$$\underline{b}_k = \underline{b}_{k-1} + \underline{k}_k(y_k - \underline{u}_k^T \hat{\underline{b}}_{k-1}) \tag{2.1}$$

Wektor $\underline{k}_k$, jest wektorem wag, z jakimi uwzględniane są poprawki dla poszczególnych składowych wektora $\hat{\underline{b}}_{k-1}$.
Wektor wag oblicza się na podstawie zależności:

$$\underline{k}_k = \underline{P}_{k-1}\, \underline{u}_k \left(1 + \underline{u}_k^T \underline{P}_{k-1}\, \underline{u}_k\right)^{-1} \tag{2.2}$$

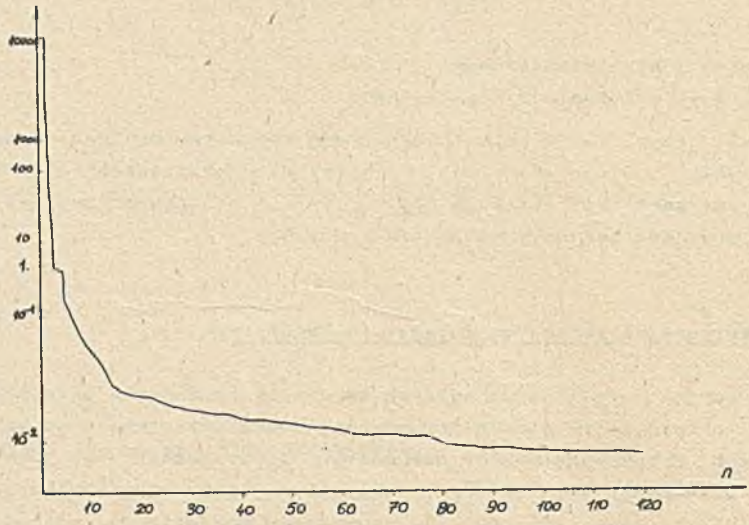
gdzie:

$$\underline{P}_k = \underline{P}_{k-1} - \underline{P}_{k-1}\, \underline{u}_k \left(1 + \underline{u}_k^T \underline{P}_{k-1}\, \underline{u}_k\right)^{-1} \underline{u}_k^T \underline{P}_{k-1} \tag{2.3}$$

Charakterystyczną cechą algorytmu RMNK jest to, że elementy macierzy $\underline{P}_k$ maleją monotonicznie ze wzrostem liczby wykonanych iteracji (rys. 3). Zapewnia to zbieżność estymat do ustalonych, chociaż niekoniecznie właściwych wartości. Duże zmiany $\hat{\underline{b}}_k$ możliwe są jedynie w początkowych iteracjach. Cecha ta wpływa korzystnie na stabilność algorytmu ale pogarsza jego zdolność nadążania za zmianami parametrów obiektu.

Dla realizacji algorytmu rekurencyjnego konieczne jest przyjęcie początkowych wartości $\hat{\underline{b}}_o$ i $\underline{P}_o$. Zwykle przyjmuje się $\hat{\underline{b}}_o = \underline{0}$ - co odpowiada całkowitemu brakowi informacji o wartości parametrów przyjętego modelu

obiektu - oraz $\underline{P}_0 = \underline{1}\,a$, gdzie a jest dużą dodatnią liczbą. To ostatnie zapewnia zbieżność algorytmu RMNK ale powoduje silne zakłócenia w początkowej fazie funkcjonowania RMNK.



Rys. 3. Przykładowy przebieg zmian elementu P(1,1)

3. Właściwości eksploatacyjne algorytmu RMNK

W stosunku do algorytmów identyfikacyjnych stawia się na ogół trzy wymagania. Żąda się aby były one szybkozbieżne, stabilne, dokładne. Z badań eksploatacyjnych przeprowadzonych przez autorkę oraz z danych literaturowych wynika, że:

1. Algorytm RMNK jest stosunkowo wolnozbieżny. Estymaty parametrów po początkowych silnych odchyleniach od swoich wartości oczekiwanych dążą w przybliżeniu aperiodycznie do wartości ustalonych.
2. Szybkość dochodzenia do wartości ustalonych poszczególnych parametrów jest różna. Najszybsza zbieżność charakteryzuje parametr b_0.
3. Ze wzrostem wymiaru wektora parametrów $\underline{b}$ wzrasta ilość iteracji potrzebnych do uzyskania przez wektor $\underline{b}$ wartości ustalonej. I tak przykładowo dla obiektów zakłócanych sygnałem pseudolosowym o wartości średniej równej zero i wariancji równej jeden, autorka uzyskała zbieżność dla estymowanych:

   a) dwóch parametrów po ok. 100 iteracjach,
   b) trzech parametrów po ok. 400 iteracjach,
   c) czterech parametrów po ok. 1000 iteracjach,
   d) sześciu parametrów po ok. 2000 iteracji.

4. Na szybkość zbieżności algorytmu ma wpływ poziom zakłóceń (stosunek maksymalnego zakłócenia do maksymalnej wartości sygnału) [7].
5. Algorytm jest wystarczająco stabilny ale utylizacja estymat jest możliwa dopiero po kilku, do kilkunastu początkowych iteracjach.
6. Przy skorelowanym zakłóceniu oddziałującym na badany obiekt, estymaty parametrów modelu dążą do wartości różnych od rzeczywistych wartości parametrów [2],[3],[6]. Dyskwalifikuje to zastosowanie algorytmu RMNK dla zakłóceń kolorowych, gdy celem stosowania RMNK jest znalezienie właściwych estymat parametrów modelu.

Z wymienionych wyżej cech RMNK najpoważniejsze konsekwencje w praktyce ma jego wolna zbieżność. Zakładając, że w przypadku identyfikacji modeli dynamicznych pomiaru wielkości wejściowych dokonuje się co 1 min, już po 4 nieznanych parametrach na uzyskanie końcowego efektu trzeba czekać ponad 16 godzin.
Ze wzrostem złożoności modelu czas oczekiwania na końcowy efekt identyfikacji wydłuża się. Powoduje to, że algorytm RMNK do wielu praktycznych zastosowań jest nieprzydatny.

4. Badanie możliwości przyspieszenia zbieżności algorytmu RMNK

Przyczyn wolnej zbieżności algorytmu RMNK upatrywano:

a) w konieczności wprowadzenia warunków początkowych $\hat{\underline{b}}_o = \underline{0}$ dla zainicjowania algorytmu,
b) w dużym wpływie historycznych wyników pomiarów przy malejącym wpływie pomiarów aktualnych na wynik estymacji.

Ponieważ wolna zbieżność algorytmu RMNK stanowi poważne ograniczenie dla praktycznych zastosowań metody, podjęto próby przyspieszenia zbieżności oparte o następujące hipotezy:

1. Wprowadzenie apriorycznej informacji o parametrach obiektu w postaci początkowej wartości wektora $\hat{\underline{b}}_o$ powinno przyspieszyć zbieżność $\hat{\underline{b}}$ do $E\{\hat{\underline{b}}\}$, jeżeli przyczyną wolnej zbieżności jest nieznajomość warunków początkowych $\underline{b}_o$.
2. Odnawianie algorytmu RMNK co n iteracji,tak, aby punkt startowy w n+1 iteracji przyjmowany był jako:

$$\hat{\underline{b}}_o = \hat{\underline{b}}_n,$$

$$\underline{P}_o = \underline{1}a,$$

przy odpowiednim doborze n powinno przyspieszyć zbieżność $\hat{\underline{b}}$ do $E\{\hat{\underline{b}}\}$, poprzez zwiększenie średniej wartości wektora wag $\underline{k}_k$.

3. Zastosowanie wygaszenia zmiennych [4],[5] powinno przyspieszyć zbieżność metody przez zwiększenie wpływu aktualnych pomiarów na wyniki estymacji.

4. Rezygnacja z dużych wartości początkowych wyrazów macierzy $\underline{P}_o$ powinna polepszyć zachowanie się algorytmu dla początkowych iteracji. Szczególnie istotną powinna być ta inowacja przy możliwości korzystania z informacji aprioryczych o parametrach obiektu.

Sprawdzenie przedstawionych hipotez przeprowadzono dla obiektu

$$y(k) = 2.1\ y(k-1) - 1.6\ y(k-2) + 0.4\ y(k-3) + e(k)$$

zakłócanego sygnałem pseudoprzypadkowym o zerowej wartości średniej i wariancji równej jeden. Dla tego obiektu dokonywano estymacji parametrów za pomocą algorytmu RMNK pracującego przez 200 iteracji.

Dla sprawdzenia hipotezy 1 uruchamiano algorytm RMNK, startując z warunków początkowych:

a) $\hat{\underline{b}}_o = \underline{0}$,

b) $\hat{\underline{b}}_o = \underline{b}$,

c) $\hat{\underline{b}}_o = 0.5\ \underline{b}$.

Początkową wartość $\underline{P}_o$ zgodnie z zaleceniami literaturowymi [2] przyjęto jako $\underline{1}\ 10^5$.

Wprowadzenie przybliżonych - pkt c i dokładnych - pkt b wartości parametrów wektora $\underline{b}$ dla punktu startowego, nie wpłynęło na przyspieszenie zbieżności algorytmu RMNK. Począwszy od piątej iteracji wartości estymowanych parametrów dla przypadków b) i c) pokrywały się z wartościami estymowanych parametrów dla pkt a).

Dla sprawdzenia hipotezy 2 uruchomiono algorytm RMNK, który pracował przez 200 iteracji,

a) w sposób ciągły,

b) po 100 iteracjach został odnowiony zgodnie z zasadą przyjętą w hipotezie 2

c) co 40 iteracji zostawał odnawiany,

d) co 25 iteracji zostawał odnawiany,

e) co 10 iteracji zostawał odnawiany,

f) co 5 iteracji zostawał odnawiany.

Dla rozważanych sześciu przypadków porównywane były wielkości:

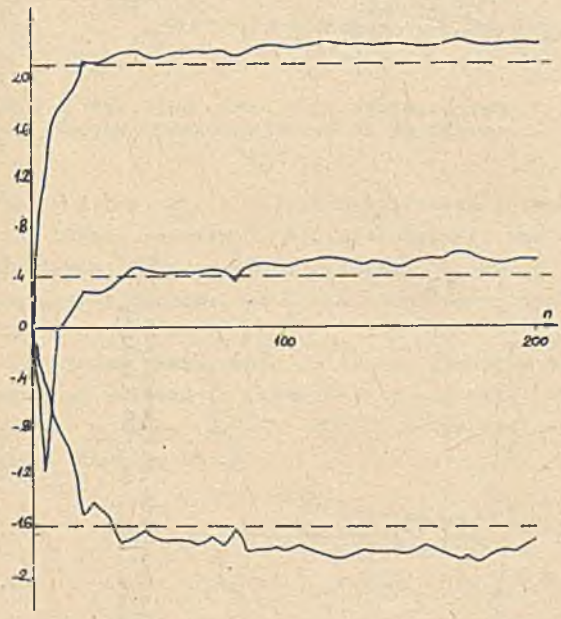
$$S = \sqrt{\frac{1}{200} \sum_{i=1}^{200} (y(i) - \hat{y}(i))^2}$$

$$S_b = \sqrt{\frac{1}{200} \sum_{i=1}^{200} \sum_{k=1}^{3} (b_k(i) - \hat{b}_k(i))^2}$$

Wyniki przedstawione są w tablicy 1.

Tablica 1

	S	S_b
a)	0.913	0.239
b)	0.881	0.429
c)	0.773	2.17
d)	0.699	3.022
e)	0.538	3.510
f)	0.307	4.836

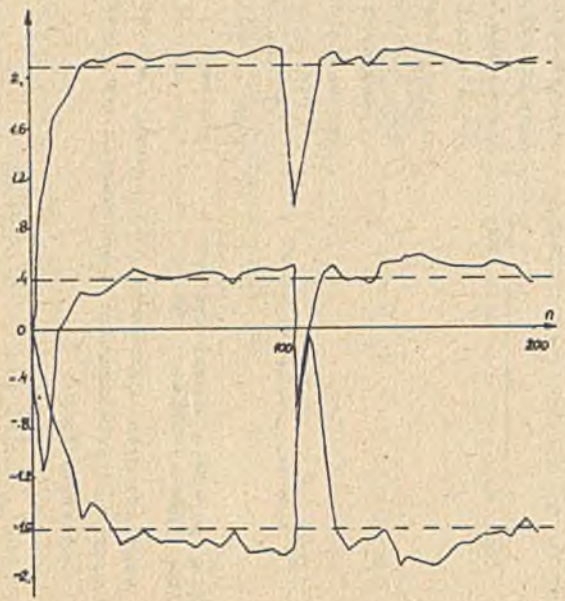


Rys. 4. Przebieg parametrów estymowanych przy pomocy RMNK

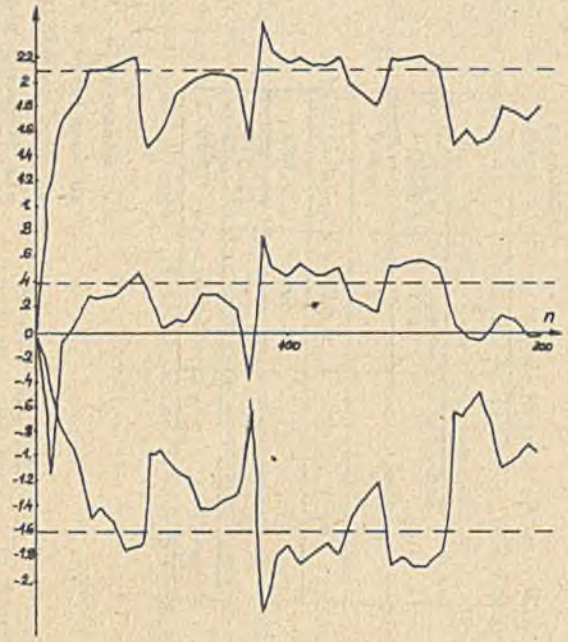
Odnawianie algorytmu rekurencyjnego w trakcie obliczeń powoduje zmniejszenie średniej sumy kwadratów błędów wyjścia tym znaczniejsze, im częściej odnawiany jest podstawowy algorytm. Odwrotnie zachowują się parametry. Ze wzrostem częstości odnawiania algorytmu zanika aperiodyczny charakter dochodzenia parametrów do ich wartości oczekiwanej, zwiększają się przeregulowania i wzrasta suma kwadratów odchyłek estymowanych parametrów od ich wartości oczekiwanych.

Dla sprawdzenia hipotezy 3 uruchomiono algorytm RMNK ze współczynnikiem wygaszania $\varrho = 0.8, 0.95, 0.98$.

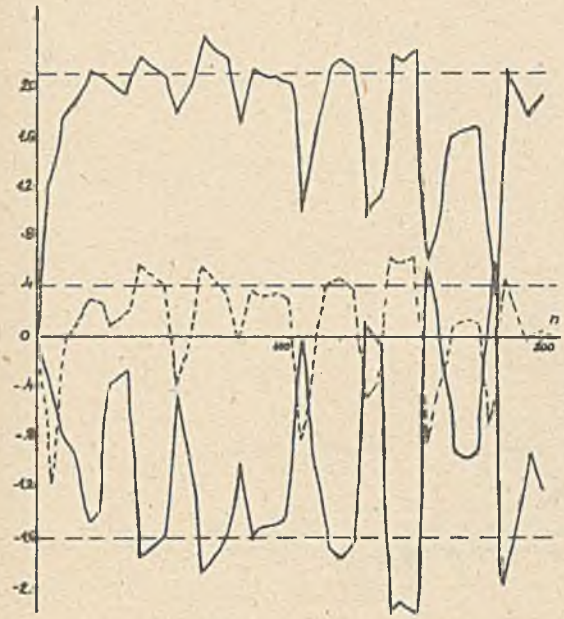
Wprowadzenie współczynnika wygaszania $\varrho < 1$ powoduje wzrost odchyłek estymowanych parametrów od ich wartości oczekiwanych tym większy, im mniejsza jest wartość współczynnika ϱ.



Rys. 5. Przebieg parametrów estymowanych przy pomocy RMNK odnawianej co 100 iteracji



Rys. 6. Przebieg parametrów estymowanych przy pomocy RMNK odnawianej co 40 iteracji



Rys. 7. Przebieg parametrów estymowanych przy pomocy RMNK odnawianej co 25 iteracji

Końcowy efekt wygaszania zmiennych ma ten sam charakter co efekt uzyskany z odnawiania algorytmu RMNK, gdyż w istocie w obu przypadkach dokonuje się tego samego zabiegu polegającego na zwiększeniu średniej wartości współczynnika wagowego $\underline{k}_k$.

Dla sprawdzenia hipotezy 4 uruchomiono algorytm startując z punktu początkowego

$$\hat{\underline{b}}_o = \underline{0}; \quad \underline{P}_o = \underline{1}\ 10^5$$

$$\hat{\underline{b}}_o = \underline{b}; \quad \underline{P}_o = \underline{1}\ 10^5$$

$$\hat{\underline{b}}_o = \underline{b}; \quad \underline{P}_o = \underline{1}\ 0.002$$

$$\hat{\underline{b}}_o = \underline{b}+\Delta\underline{b}; \quad \underline{P}_o = \underline{1}\ 0.002$$

Uzyskane rezultaty przedstawione są na rys. 8a.

W przypadku dokładnej informacji o wartości parametrów estymowanych zastosowanie macierzy $\underline{P}_o$ o małych wartościach elementów na przekątnej głównej, daje lepsze rezultaty, niż zastosowanie macierzy o dużych elementach diagonalnych. Ze wzrostem odległości $\hat{\underline{b}}_o$ od $\underline{b}$ efektywność RMNK z dużymi początkowymi wartościami elementów diagonalnych macierzy $\underline{P}_o$ wyraźnie rośnie.

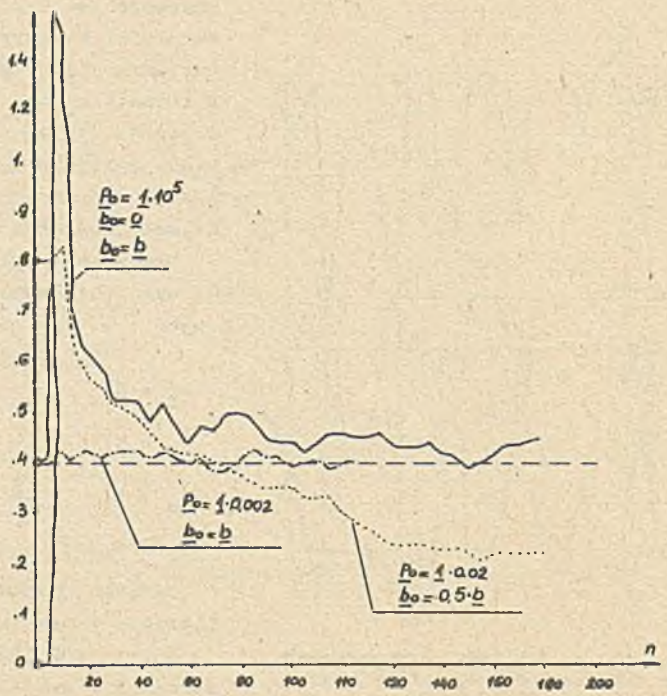
Ponieważ praktycznie inicjując algorytm nie posiadamy informacji o wartościach wektora $\underline{b}$ uruchamiano algorytm RMNK startując z punktu początkowego $\hat{\underline{b}}_o = \underline{0}$, $\underline{P}_o = \underline{1}\ 10^5$, a po pięćdziesięciu iteracjach zmieniono punkt startowy na:

$$\hat{\underline{b}}_o = \hat{\underline{b}}_{50}, \quad \underline{P}_o = 0.002\ \underline{1}$$

$$\hat{\underline{b}}_o = \hat{\underline{b}}_{50}, \quad \underline{P}_o = 0.2\ \underline{1}$$

Uzyskane wyniki przedstawia rys. 8b

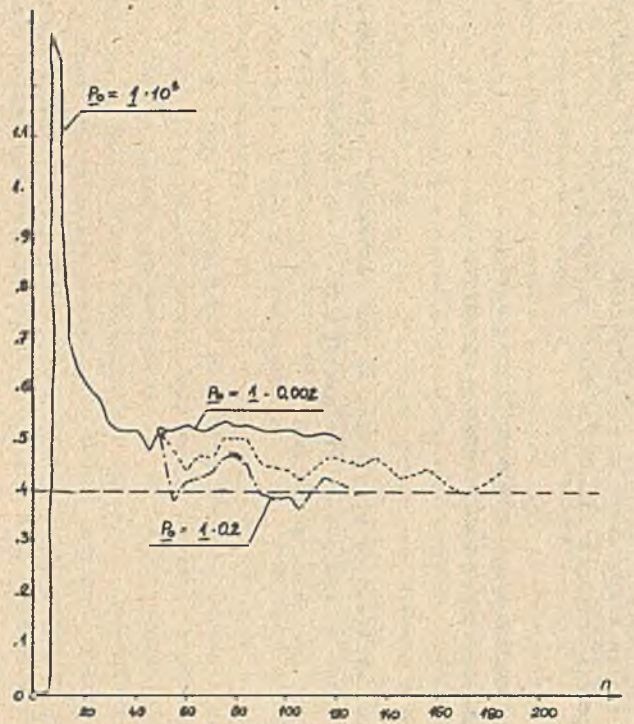
Rysunek 8c przedstawia wyniki sekwencyjnego odnawiania algorytmu RMNK co 25 iteracji z różnymi początkowymi wartościami macierzy $\underline{P}_o$.



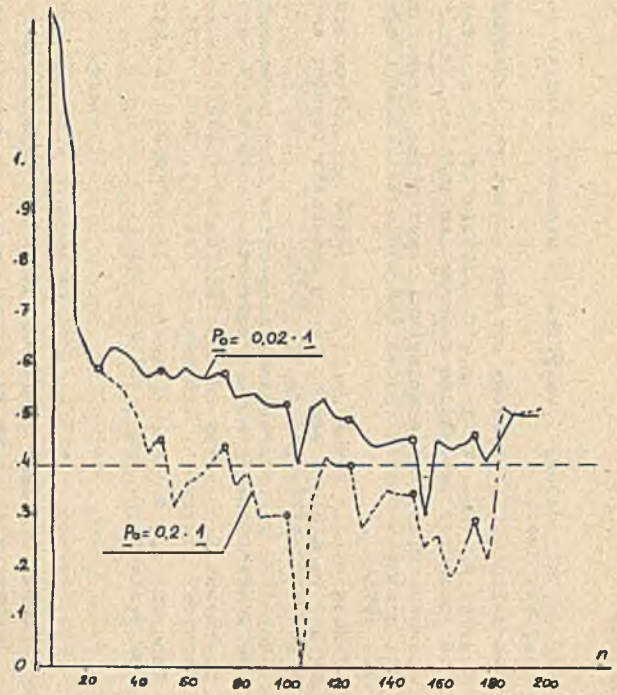
Rys. 8a. Zachowanie się algorytmu RMNK dla różnych warunków początkowych

Przeprowadzone badania pozwalają stwierdzić, że, gdy celem stosowania algorytmu RMNK (przy braku apriorycznej informacji o parametrach) jest uzyskanie możliwie dokładnych estymat parametrów $\underline{b}$, najefektywniejszy pozostaje klasyczny algorytm RMNK ze współczynnikiem wygaszania = 1, startujący z dowolnych warunków początkowych (np. $\hat{\underline{b}}_o = 0$ i $\underline{P}_o = \underline{1}\, a$), gdzie a jest odpowiednio dużą liczbą.

Zabiegi wykorzystane dla przyspieszenia algorytmu RMNK, pomimo niewielkiej ich efektywności w ogólnym przypadku stosowania RMNK, mogą znaleźć zastosowanie do identyfikacji modeli obiektów o zmieniających się parametrach. Pojawia się w tym przypadku znany problem dobrania odpowiedniego współczynnika wygaszania ϱ, odpowiedniej częstotliwości odnawiania algorytmu RMNK lub odpowiedniej wartości $\underline{P}_o$. Jednak bez pobieżnej choćby informacji o zakłóceniach i wartościach szukanych parametrów nie można wybrać a priori najlepszego sposobu uczulenia RMNK na zmianę estymowanych parametrów.



Rys. 8b. Zachowanie się algorytmu RMNK odnowionego po 50 iteracjach dla różnych wartości P_0



Rys. 8c. Zachowanie się algorytmu RMNK odnawianego co 25 iteracji dla różnych wartości P_0

LITERATURA

[1] Mańczak K.: Metody identyfikacji wielowymiarowych obiektów sterowania, WNT 1971.

[2] Niederliński A.: Systemy cyfrowe automatyki przemysłowej. WNT, 1977.

[3] Isermann R., Baur U. i in.: Comparison of six on-line identification and parameter estimation methods, Automatica, Vol. 10, No 1, 1974.

[4] Scheurer H.G.: Ein adaptatives, explizites Parameterschatz-verfahren mit geringem Speicherplatz und Rechenzeitbedarf Regelungstechnik Heft 12, 1975.

[5] Ziółkowski A.: Wyznaczanie na bieżąco parametrów modelu na podstawie kryterium minimum ważonej sumy kwadratów błędów. Archiwum Automatyki i Telemechaniki, zeszyt II, 1974.

[6] Söderström T.: Convergence properties of the Generalised least squares identification method Automatica Vol. 10, No 6, 1974.

[7] Gustavson I.: Survey of application of identification in chemical and physical processes. Automatica, Vol. 11, No 1, 1975.

[8] Nahorski Z.: O jednoznaczności identyfikacji obiektów liniowych z czasem dyskretnym. Prace VII KKA tom 1.

О ВОЗМОЖНОСТИ ПОВЫШЕНИЯ СХОДИМОСТИ АЛГОРИТМА ИТЕРАТИВНОГО МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Р е з ю м е

Оговорено эксплуатационные свойства алгоритма оценки параметров обоснованного на итеративном методе наименьших квадратов. Рассмотрено возможности повышения скорости сходимости этого алгоритма.

SOME REMARKS ABOUT THE CONVERGENCE RATE OF THE RECURSIVE LEAST SQUARES ALGORITHM

S u m m a r y

The utilisation properties of the recursive least squares algorithm are discussed in the peper. The possibilities of increasing of the convergence rate are being concerned.