Noname manuscript No.
(will be inserted by the editor)

Podejście parametryczne w identyfikacji systemów

Rafał Lorek

Wrocław, 2018

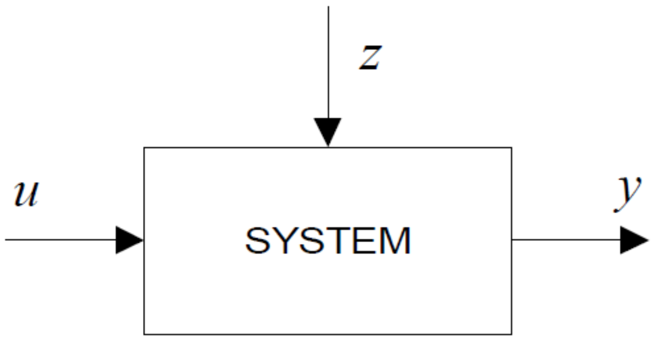
Streszczenie Dokument opisuje wybrane zagadnienia z dziedziny identyfikacji systemów.

Address(es) of author(s) should be given

Spis treści

1 Czym jest identyfikacja

Identyfikacja systemów to rozpoznawanie (sporządzenie opisu matematycznego) właściwości statycznych i dynamicznych elementów i układów automatyki. Identyfikacja oznacza znalezienie zależności między wejściem a wyjściem (dla elementu automatyki, obiektu, układu regulacji) na podstawie danych doświadczalnych. Po poddaniu obiektu (procesu) szeregowi doświadczeń dobiera się bowiem parametry modelu w taki sposób, aby pasował on do danych doświadczalnych. Identyfikacja odgrywa zasadniczą rolę w odniesieniu do obiektów i procesów regulacji, gdyż umożliwia poprawne nastrojenie układu regulacji automatycznej. W czasie identyfikacji określane są bowiem wartości parametrów modelu obiektu (procesu), które wykorzystuje się następnie w doborze nastaw regulatora sterującego rzeczywistym obiektem (procesem).



Rysunek 1 Istota identyfikacji

2 Etapy identyfikacji parametrycznej:

1. Przygotowanie eksperymentu identyfikacyjnego, generacja pobudzeń.

2. Przeprowadzenie eksperymentu identyfikacyjnego, zebranie pomiarów.

3. Wstępne przetwarzanie danych pomiarowych (np. eliminacja błędów tzw. grubych).

4. Wybór klasy dopuszczalnych modeli:
   - deterministyczne lub stochastyczne
   - ciągłe lub dyskretne
   - liniowe lub nieliniowe
   - stacjonarne lub niestacjonarne

5. Wybór typu modelu z wybranej klasy.

   W każdej klasie modeli istnieją różne ich typy (np. w klasie dyskretnych modeli stochastycznych można wyróżnić modele ARX, MAX, ARMAX, ARIX). W tym kroku należy dokonać wyboru na podstawie wstępnej analizy obiektu lub zarejestrowanych sygnałów.

6. Wybrór struktury modelu na podstawie przeglądu struktur danego typu.

7. Estymacja parametrów danego modelu.

   W tym kroku podejmuje się decyzję na temat algorytmu estymacji, który pozwoli na wyznaczenie parametrów wcześneij wybranego modelu.

8. Weryfikacja modelu

   Etap ten kończy pojedyńczą iterację procesu identyfikacji. W tym kroku należy rozstrzygnąć, czy wynik identyfikacji jest zadowalający.
   W tym celu można dokonać:
   - porównania sygnału wyjściowego modelu z sygnałem rzeczywistym (najlepiej dla innego zbioru danych)
   - wyznaczenia błędu predykcji jednokrokowej i określenia jego cech
   - weryfikacji, czy model jest nadparametryzowany (czy posiada zbyt bogatą strukturę)
   - weryfikacji innych cech modelu, decydujące o jego przydatności (np. stabilność czy odwracalność)

Uzyskanie niezadowalających wyników w trakcie weryfikacji modelu powoduje konieczność powtórzenia niektórych etapów identyfikacji. Zazwyczaj najpierw powtarza się etapy 6-8, gdy to nie przynosi efektu, należy powtórzyć kroki począwszy od 4.
W odniesieniu do rzeczywistości powtórzenie kroków 1-2 wiąże się z dużymi kosztami. Z tego powodu należy dokonać wszelkich starań aby przygotowanie eksperymentu było miżliwie najlepsze (min. poprzez staranny wybór sygnałów pobudzających).

3 Sposób wybierania metody identyfikacji

– ze względu na zakres przyjętych założeń o systemie

  Metody parametryczne, gdy zakładamy, że postacie funkcji opisujących system są nam znane z dokładnoącią do parametrów.
  Metody nieparametryczne, gdy system traktujemy jako "czarną skrzynkę", tj. zakładamy jedynie istnienie adekwatnego opisu w określonej klasie charakterystyk

– ze względu na charakter sygnałów występujących w systemie
  (problemy deterministyczne; problemy probabilistyczne)

– ze względu na rodzaj charakterystyki opisującej system
  (identyfikacja systemów liniowych; identyfikacja systemów nieliniowych)

– ze względu na zachowanie się systemów w funkcji czasu lub sposobu pomiaru sygnałów w czasie (identyfikacja systemów z czasem ciągłym; identyfikacja systemów z czasem dyskretnym) obecnie częściej stosowana ze względu na rozwój cyfrowej techniki obliczeniowej

– ze względu na własności badanego systemu (identyfikacja systemów statycznych; identyfikacja systemów dynamicznych)

– ze względu na strukturę systemu (identyfikacja systemów (obiektów) prostych (jednoelementowych); identyfikacja systemów o złożonej strukturze, tzw. blokowozorientowanych (tj. zbioru wzajemnie połączonych i współzależnych podsystemów prostych - liniowych obiektów dynamicznych i nieliniowych obiektów statycznych). W ramach tej kategorii można dokonać dalszego podziału na identyfikację systemów o strukturach otwartych i zamkniętych (zawierających sprzężenie zwrotne).

Podstawową przesłanką przy wyborze algorytmu identyfikacji powinna być informacja wstępna o systemie, tj. o postaci jego charakterystyk i rodzaju zakłóceń. Czasami informacja wstępna jaką dysponujemy może być bardzo dokładna (np. wiemy że obiekt jest liniowy i znamy rozkład występujących w nim zakłóceń), często jednak nasza wiedza ogranicza się do ogólnych informacji jakościowych (np. o ograniczoności wariancji zakłóceń, czy ciągłości charakterystyki obiektu). W zależności od posiadanej informacji decydujemy się na odpowiednią metodę identyfikacji. Wyróżniamy dwie wzajemnie uzupełniające się grupy metod – parametryczne i nieparametryczne. W metodach parametrycznych zakłada się znajomość postaci funkcji opisujących system (np. liniowe, wykładnicze,wielomianowe, itp.), tj. ich charakterystyk (opisów) z dokładnością do parametrów.

4 Metody parametryczne

Polegają one na znajdowaniu prawdziwego wektora parametrów systemu, lub takiego wektora współczynników dla którego założony model najlepiej przybliża rzeczywisty system. Algorytmy takie działają na zasadzie minimalizacji pewnego wskaznika niedokładności modelu (jakości identyfikacji) ze względu na poszukiwany wektor parametrów. Znalezienie takiego minimum może stanowić skomplikowane zadanie obliczeniowe. Zaletą metod parametrycznych jest duża praktyczna szybkość zbieżności, niezależna od wymiarowości problemu i rodzaju charakterystyk oraz ich prostota z inżynierskiego punktu widzenia. W niektórych przypadkach możliwa jest zadowalająco dokładna identyfikacja systemu, nawet na podstawie kilku pomiarów. Niebezpieczeństwem podejścia parametrycznego jest natomiast możliwość wystąpienia błędu systematycznego jaki może się pojawić przy przyjęciu niewłaściwego modelu (założeniu błędnej postaci charakterystyki obiektu).

Najpopularniejsze metody parametryczne:

- metoda najmniejszych kwadratów (ang. least square)
- metoda zmiennych instrumentalnych (ang. instrumental variables)
- metoda największej wiarygodności
- metoda największego prawdopodobieństwa

5 Wybrane parametryczne metody identyfikacji systemów

5.1 Metoda Najmniejszych Kwadratów

W statystyce wykorzystuje się ją do estymacji i wyznaczania linii trendu na podstawie zbioru danych w postaci par liczb. Najczęściej jest stosowana przy regresji liniowej, ale może też być stosowana do statystycznego wyznaczania parametrów nieliniowych linii trendu.

Dla liniowego obiektu statycznego opisanego równaniem

$$y_i = x_i^T b^* + z_i$$

gdzie y_i, x_i oraz z_i oznaczają odpowiednio wyjście, wektor wejściowy oraz zakłócenie w i-tym pomiarze $(i = 1, ..., N)$ oceną prawdziwego wektora parametrówb^* według metody LS jest wektor $b_N^{(LS)}$ postaci:

$$b_N^{(LS)} = (X_N^T X_N)^{-1} X_N^T Y_N$$

gdzie

$$X_N = (x_1, x_2, \dots, x_N)^T$$
$$Y_N = (y_1, y_2, \dots, y_N)^T$$
$$Z_N = (z_1, z_2, \dots, z_N)^T$$

Estymator LS jest estymatorem liniowym ze względu na pomiary wyjść. Oznaczając

$$L_N = (X_N^T X_N)^{-1} X_N^T$$

możemy zapisać jako

$$b_N^{(LS)} = L_N Y_N$$

przy czym

$$Y_N = X_N b^* + Z_N$$

Przy następujących założeniach o losowych zakłóceniach i wejściach:
- ciągi $\{x_i\}$ i $\{z_i\}$ są typu iid i nie zależą od siebie
- $E_z = 0$
- $varz < \infty$
- $Exx^T < \infty$

estymator $b_N^{(LS)}$ jest estymatorem nieobciążonym parametrów b^*, tzn.

$$b_N^{(LS)} \to b^*$$

z prawdopodobieństwem 1, gdy $N \to \infty$

W liniowym dyskretnym obiekcie dynamicznym wartość wyjścia zależy również od wartości wejść i wyjść w chwilach poprzednich. Obiekt taki (typu ARX) opisuje równanie różnicowe

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} \dots - a_R y_{k-R} + b_0 x_k + b_1 x_{k-1} + \dots + b_s x_s + z_k$$

gdzie $\{y_k\}$ jest procesem wyjściowy, $\{x_k\}$ - procesem wyjściowym, $\{z_k\}$ - zakłóceniem. natomiast R i S określają rząd autoregresji i ruchomej średniej w opisie obiektu.
Definiując uogólniony wektor wejść

$$\varphi = (-y_{k-1}, \ldots, -y_{k-R}, x_k, x_{k-1}, \ldots, x_{k-S})^T$$

oraz wektor prawdziwych wartości parametrów obiektu

$$p^* = (a_1, \ldots, a_R, b_0, b_1, \ldots, b_s)^T$$

równanie obiektu dynamicznego można przedstawić jako regresję liniową, w postaci podobnej do opisu obiektu statycznego, tj.:

$$y_k = \varphi_k^T p^* + z_k$$

Podobnie możliwe jest skonstruowanie równania macierzowego opisującego zestaw N pomiarów. Oznaczając:

$$Y_N = (y_1, y_2, \ldots, y_N)^T$$
$$Z_N = (z_1, z_2, \ldots, z_N)^T$$
$$\Phi_N = (\varphi_1, \varphi_2, \ldots, \varphi_N)^T$$

otrzymjemy

$$Y_N + \Phi_N p^* + Z_N$$

Ze względu na powyższe analogie, estymator LS dla liniowego obiektu dynamicznego ma postać:

$$p_N^{LS} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N$$

W praktycznych przypadkach zakłócenie z_k występujące w opisie systemu nie jest w ogólności białym szumem, jak miało to miejsce w przypadku identyfikacji systemu statycznego, ale procesem skorelowanym. Wartość z_k zależy od zakłóceń w chwilach poprzednich. Proces z_k należy zatem traktować ogólnie jako stacjonarny (w stanie ustalonym) szum kolorowy. Macierz kowariancji zakłóceń nie jest wtedy diagonalna. Pierwsze R elementów uogólnionego wektora wejść φ_k może być skorelowane z zakłóceniem z_k. Z tego powodu estymator najmniejszych kwadratów dla liniowego obiektu dynamicznego jest w ogólności obciążony, również asymptotycznie.
Oznaczając przez

$$\sum = E Z_N Z_N^T$$

macierz kowariancji zakłóceń (z założenia $E_{z_k} = 0$) i zakładając, że macież Σ jest dodatnio określona (ozn. $\Sigma > 0$) możemy zapisać

$$\sum = P_\Sigma P_\Sigma^T$$

gdzie P_Σ jest macierzą nieosobliwą. Mnożąc lewostronnie równanie przez macież P_{sigma}^{-1} i oznaczając

$$\overline{Y_N} = P_\Sigma^{-1} Y_N$$
$$\overline{\Phi_N} = P_\Sigma^{-1} \Phi_N$$
$$\overline{Z_N} = P_\Sigma^{-1} Z_N$$

otrzymujemy

$$\overline{Y_N} = \overline{\Phi_N} + \overline{Z_N}$$

gdzie po przekształceniu, elementy wezktora zakłóceń wektora Z_N w powyższym równaniu tworzą biały szum, bowiem

$$cov(\overline{ZN}) = E\overline{Z_N ZN}^T = P_\Sigma^{-1} E Z_N Z_N^T P_\Sigma^{T-1} = I$$

Macierz P_Σ^{-1} pełni rolę filtru wybielającego zakłócenia. W odniesieniu do powyższego równania w celu oszacowania wektora p^* możena zastosować estymator LS w postaci

$$\overline{p_n^{(LS)}} = (\overline{\Phi_N^T \Phi_N})^{-1} (\overline{\phi_N^T Y_N})$$

Taki estymator posiada wszystkie własności estymatora LS dla systemów statycznych, wymaga jednak dokładej zanjomości budowy korelacyjnej zakłóceń, tj. macierzy $\sum$, której z reguły nie znamy. Istnieją różne metody ominięcia problemu skorelowania dla różnych specyficznych modeli zakłóceń. W praktyce najczęściej przyjmuje się następujące modele zakłóceń, uwzględniające różne rodzaje skorelowania. Powstają one w rezultacie odpowiedniej filtracji białego szumu, tj. procesu ϵ_k typu iid takiego, że $E\epsilon_k = 0$ oraz $var\epsilon_k < \infty$
Znajomość budowy korelacyjnej zakłóceń (macierzy $\sum$) umożliwia filtrację wybielającą zakłóceń, która sprowadza problem do zadania identyfikacji z białym szumem. W praktyce struktura korelacyjna zakłóceń nie jest jednak znana i zazwyczaj występuje uciążliwa konieczność jednoczesnego identyfikowania parametrów toru zakłóceń. Uniwersalną metodą postępowania, stosowaną dla systemów liniowych niezależnie od struktury korelacyjnej zakłóceń jest metoda zmiennych instrumentalnych.

5.2 Metoda Zmiennych Instrumentalnych (IV)

Gdy w zadaniu identyfikacji liniowego obiektu dynamicznego występuje skorelowanie zakłóceń można zastosować metodę zmiennych instrumentalnych ($ang.IV - instrumentalvariablemethod$). Polega ona na zastąpieniu estymatora LS estymatorem postaci:

$$p_N^{(IV)} = (\Psi_N^T \Phi_N)^{-1} \Psi_N^T Y_N$$

gdzie

$$\Psi_N = (\psi_1, \psi 2, \dots, \psi_N)^T$$

jest odpowiednią dodatkową macierzą oraz

$$\psi = (\psi_{k,1}, \dots, \psi_{k,R}, \psi_{k,R+1}, \dots, \psi_{k,R+S+1})^T$$

jest wektorem o tym samym wymiarze co wektor φ_k. Macierz Φ_N jest zatem macierzą o strukturze i wymiarach zgodnych z macierzą Φ_N

$$dim \Psi_N = dim \Phi_N$$

Zawiera ona tzw. zmienne instrumentalne $\psi_{k,i}$ (nstrumenty). Stąd jest nazywana macierzą zmiennych instrumentalnych. Procedura identyfikacji systemu w oparciu o metodę zmiennych instrumentalnych wymaga zatem generacji macierzy Ψ_N. Aby ustalić jakie własności powinny spełniać zmienne instrumentalne, po to aby estymator IV był zgodny, mnożymy lewostronnie równanie pomiarów przez Ψ_N^T

$$\Psi_N^T Y_N = \Psi_N^T \Phi_N p_N^{(IV)}$$

Po podstawieniach i przekształceniach otrzymujemy:

$$(\frac{1}{N} \Psi_N^T \Phi_N)(p_N^{IV} - p^*) = (\frac{1}{N} \Psi_N^T Z_N)$$

Wynikają stąd dwa podstawowe postulaty jakie równocześnie muszą spełniać zmienne instrumentalne.

$PostulatI$

Instrumenty powinny być takie, ze istnieje granica $P \lim_{N \to \infty} (\frac{1}{N} \Psi_N^T \Phi_N)$

$$P \lim_{N \to \infty} (\tfrac{1}{N} \Psi_N^T \Phi_N) = E_{\psi_k \varphi_k^T}$$

i jest ona macierzą nieosobliwą

$$\det E \psi_k \phi_k^T \neq 0$$

$PostulatII$

Instrumenty powinny być takie, że

$$P \lim_{N \to \infty} (\tfrac{1}{N} \Psi_N^T \Phi_N) = E_{\psi_k \phi_k^T}$$

oraz

$$E \psi_k z_k = 0$$

Dla skuteczności metody wystarczy, że macierz graniczna będzie dowolną macierzą nieosobliwą, a wektor graniczny wektorem zerowym. Do najbardziej popularnych metod generacji macierzy Ψ_N należą:

- przypisanie zmiennym pomocniczym wartości wejść w chwilach poprzednich
$$\psi_k = (x_{k-1} \dots, x_{k-R}, x_{k-R-1}, \dots, x_{k-R-S-1})$$
- zastosowanie dowolnej filtracjiliniowej sygnału wejsciowego
$$\psi_{k,j} = \sum_{i=0}^{\infty} h_{j,i} x_{k-i}$$
- interacyjne wyznaczanie elementów wektora ψ_k

6 Wady i zalety metod parametrycznych w identyfikacji systemów

- Efektywność metody IV dla systemów liniowych jest silnie wrażliwa na zastosowane zmienne instrumentalne.
- Metoda najmniejszych kwadratów dostosowuje się bowiem do punktów najbardziej oddalonych od średniej, które mogą wprowadzić największy błąd. Jeśli mamy w danych pojedynczą zakłócającą obserwację bardzo oddaloną od reszty, przyciągnie ona do siebie linię trendu. Takie zjawisko jest niestety częste w realnych danych, nie należy więc stosować metody najmniejszych kwadratów bez sprawdzenia (choćby na wykresie rozrzutu) braku elementów odstających i ich usunięcia.
- Zadania identyfikacji oparte na metodach LS i IV prowadzą generalnie do problemów słabo uwarunkowanych nuerycznie (tzn. odwracana macierz ma wyznacznik bliski zeru). Powoduje to dużą zależność wyników od błędów zaokrągleń danych wejściowych i zastosowanego algorytmuobliczeniowego. Najczęściej stosuje się techniki odbić Householdera, obrotów Givensa lub wersje rekurencyjne algorytmów.
- Wadą LS jest konieczność zastosowania odwrócenia macierzy, co w niektórych przypadkach może okazać się niemożliwe. Aby tego uniknąć stosuje się rozkład SVD ($singularvaluedecomposition$) według wartości szczególnych. Pozwala to na rozkład dowolnej macierzy pierwotnej (gdy macierz posiada mniej wartości własnych niż wynosi długość przekątnej to wstawiamy 0). Dla systemów nieliniowych (wtakich, których charakterystyka zależy nieliniowo od nieznanych parametrów $[np : e^{kx}]$) stosuje się bardziej zaawansowane metody np. metaheurystyczne.

Literatura

1. Grzegorz Mzyk, Parametryczna identyfikacja systemów o złożonej strukturze (rozprawa doktorska) , Wrocław, 2002
2. Michał Marzyński, Podejście parametryczne i nieparametryczne w identyfikacji systemów, Wrocław 2013
3. K. Mańczak, Z. Nahorski, Komputerowa identyfikacja obiektów dynamicznych, PWN, Warszawa 1983