СТРУКТУРНА ІДЕНТИФІКАЦІЯ МОДЕЛЕЙ ЯК ЗАДАЧА ВІДНОВЛЕННЯ СИГНАЛУ В УМОВАХ НЕПОВНОТИ ІНФОРМАЦІЇ

В.С. Степашко,

Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій і систем

НАН та Міносвіти і науки України

Вступ

Задачу моделювання за даними спостережень, або структурної ідентифікації моделей складних об'єктів [1], за стохастичних припущень можна розглядати як задачу виділення сигналу на фоні завад [2]. У доповіді аналізується проблема вибору найкращої моделі об'єкта за короткою вибіркою даних в умовах неповноти інформації, яка характеризується: невеликим обсягом даних; невідомим характером і рівнем шуму; відсутністю інформації про склад релевантних факторів та параметри моделі. Невизначеним є також вибір ефективного методу структурної ідентифікації, який характеризується передусім критерієм оцінювання якості генерованих моделей.

Розглядається клас задач структурної ідентифікації, що так чи інакше зводяться до вибору оптимальної за заданим критерієм моделі з множини генерованих моделей, лінійних за параметрами. Вибір структури моделі з мінімальною дисперсією помилки прогнозування, або завадостійкої моделі, приймається в якості головної мети розв'язання задачі. Метою аналізу є дослідження залежності оптимального вибору моделі від можливих значень показників неповноти інформації, а також порівняльний аналіз ефективності різних критеріїв. Для цього використовується новий аналітичний апарат теорії структурної ідентифікації – метод критичних дисперсій, який можна застосовувати як для обмеженої вибірки, так і в асимптотиці.

Постановка задачі

Задано вибірку $W=(X\,|\,y)$ даних у вигляді $n\times m$ матриці X і $n\times 1$ вектора y, що містять інформацію n спостережень за m незалежними входами і одним виходом статичного об'єкта. Матриця X вважається детермінованою повного рангу $rank\,X=m$, а вектор y містить шум:

$$y=\overset{o}{y}+\xi,\quad \overset{o}{y}=Ey=X\theta_0,\quad E\xi=0,\quad E\xi\xi^T=\sigma^2 I_n\,, \tag{1}$$

де $\overset{o}{y}$ – точний (незашумлений) вихід об'єкта, θ_0 – точний (невідомий) вектор параметрів об'єкта, ξ – вектор шуму з незалежними, однаково розподіленими компонентами, E – символ математичного сподівання за всіма можливими реалізаціями вектора шуму, I_n – одинична $n\times n$ матриця, σ^2 – невідома скінченна дисперсія шуму.

Задача полягає у виборі найкращої за заданим критерієм моделі з деякої їх множини, сформованої певним генератором. В найпростішому випадку множина $\Im_m$ порівнюваних моделей містить m "вкладених" структур з послідовним ускладненням на один лінійний член:

$$\Im_m:\ \hat{y}_s=X_s\hat{\theta}_s,\quad s=\overline{1,m}\,, \tag{2}$$

де s – число оцінюваних параметрів, X_s – підматриця з s довільних (наприклад, перших) стовпців матриці X, $\hat{\theta}_s$ – МНК-оцінка на W. Вираз (2) характеризує довільний шлях ускладнення моделі, тому розгляд лише множини $\Im_m$ не впливає на загальність аналізу задачі.

Метою задачі моделювання будемо вважати вибір моделі з мінімальною *дисперсією помилки відновлення* сигналу в n точках [2,3]:

$$J(s) = \mathrm{E}\left\| \overset{o}{y} - \hat{y}_s \right\|^2 = E\left\| \overset{o}{y} - X_s \hat{\theta}_s \right\|^2 . \tag{3}$$

Цей "ідеальний" критерій характеризує якість відновлення істинного сигналу $\overset{o}{y}$ за зашумленим y з допомогою оцінки $\hat{y}_s$.

Аналогічно (3), якість прогнозу в n_F заданих точках $X_F \neq X$ за моделлю складності s (2), можна характеризувати величиною сумарної *дисперсії помилки прогнозування* $J_F(s)$ сигналу в області X_F [4]. Критерії $J(s)$ і $J_F(s)$ містять точні (не вимірювані) вектори виходу об'єкта, тобто відображають "ідеальну" мету моделювання. На практиці треба застосовувати деякі їхні оцінки.

Структури моделей, що мають складність (число параметрів)

$$s^o = arg \min_{s=\overline{1,m}} J(s) \; , \; s_F^o = \arg \min_{s=\overline{1,m}} J_F(s) \; , \tag{4}$$

називаються відповідно *J-оптимальною* (оптимальною *відновлюючою*) та J_F*-оптимальною* (оптимальною *прогнозуючою*) структурами.

Властивості дисперсії помилки моделі

Із (3) випливає вираз для $J(s)$ як функції дискретного аргументу:

$$J(s) = J^b(s) + J^v(s) = \left\| \overset{o}{y} - X_s \bar{\theta}_s \right\|^2 + \sigma^2 s \; , \tag{5}$$

де $\bar{\theta}_s = E\left[\hat{\theta}_s\right]$, індексом "$b$" (bias) позначено втрати $J^b(s)$ від зміщення структури моделі (*структурна* складова), а індексом "v" (variation) – варіацію функціоналу $J^v(s)$ через наявність шуму (*шумова* складова).

У свою чергу, формула для $J_F(s)$ набуває вигляду [4]:

$$J_F(s) = J_F^b(s) + J_F^v(s) = \left\| \overset{o}{y}_F - X_{Fs}\bar{\theta}_s \right\|^2 + \sigma^2 tr\left[\left(X_s^T X_s \right)^{-1} X_{Fs}^T X_{Fs} \right], \tag{6}$$

Характер залежності від s не очевидний, тому слід перейти до рекурентних по s співвідношень.

Позначивши $X_{s+1} = (X_s \mid x), \bar{\theta}_{s+1} = (\bar{\theta}_s^T \bar{\vartheta}_{s+1})^T$, отримаємо:

$$J(s+1) = J(s) - \bar{\vartheta}_{s+1}^2 \beta_{s+1} + \sigma^2 , \tag{7}$$

$$\bar{\vartheta}_{s+1} = \alpha_{s+1} / \beta_{s+1} , \quad \alpha_{s+1} = x^T D_s \overset{o}{y}, \quad \beta_{s+1} = x^T D_s x , \tag{8}$$

де $D_s = I_n - X_s (X_s^T X_s)^{-1} X_s^T$ – ідемпотентна матриця, а $\beta_{s+1} > 0$ для всіх $s = 0,1,...,m-1$. Тобто $J^b(s)$ як функція дискретного аргумента s строго монотонно спадає до нуля, а $J^v(s)$ лінійно зростає, тому $J(s)$ завжди має мінімум у деякій точці $s^o \in [1, m]$. У разі *збільшення* рівня (дисперсії) шуму σ^2 значення s^o *зменшується*, тобто оптимальна структура спрощується. Цю закономірність відбито на рис.1.

Відповідні рекурентні формули для $J_F(s)$ одержано в [4] і показано, що структурна складова $J_F^b(s)$ загалом немонотонно спадає до нуля, а шумова $J_F^v(s)$ строго монотонно зростає. Тобто функція $J_F(s)$, як і $J(s)$, має мінімум (глобальний) всередині інтервалу зміни дискретного аргумента $s \in [1, m]$, і цей мінімум відповідає тим меншій оптимальній складності s_F^o моделі, чим більша дисперсія шуму σ^2. Отже, поведінку функції $J_F(s)$ якісно характеризує той же рис.1.

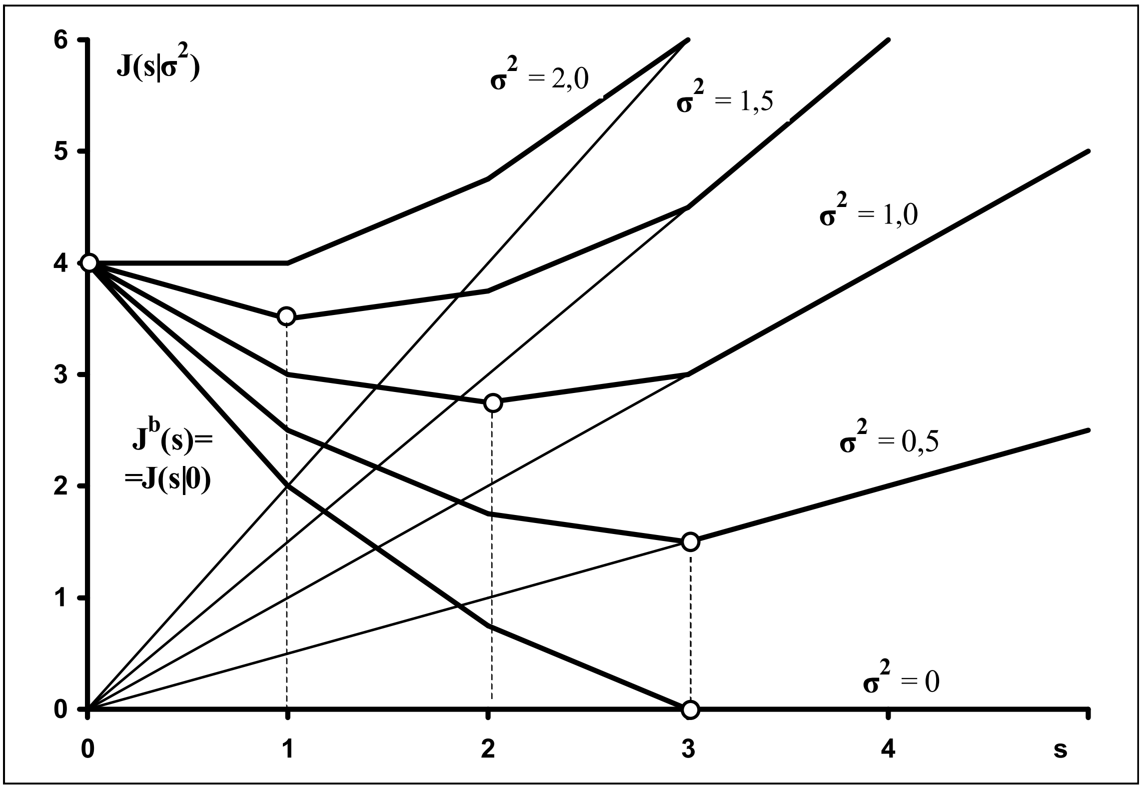

Рис.1. Дисперсія помилки моделі як функція $\boldsymbol{s}$ за різних дисперсій шуму (кружечками вказано положення мінімумів критерію $J(s)$, $s_0 = 3$, $m = 5$)

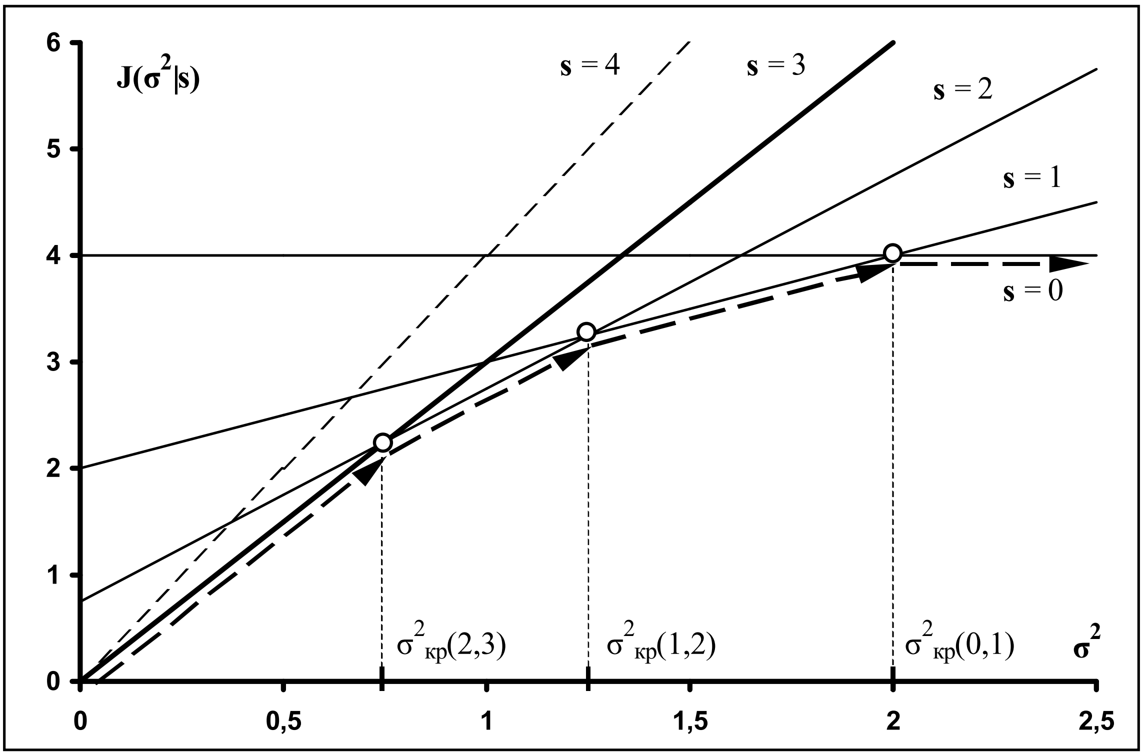

Рис.2. Дисперсія помилки моделі як функція дисперсії шуму σ^2 для структур різної складності (кружечками вказано критичні точки зміни структури моделі, яким відповідають критичні дисперсії $\sigma^2_{кр}(s,s+1)$)

Метод критичних дисперсій

Графіки на рис.1 якісно характеризують принцип завадостійкого моделювання, але не дозволяють кількісно оцінити умови оптимальності тієї чи іншої структури моделі. Тому доцільно перейти до значно інформативнішого зображення залежності дисперсії помилки моделі від дисперсії шуму σ^2 для різних структур моделей (різних s). Така залежність, як видно з (5), (6), є лінійною.

Для побудови залежностей $J(\sigma^2 \mid s)$ на осі ординат графіка (рис.2) слід спочатку відкласти ординати структурної складової $J^b(s)$, а далі провести прямі лінії, що відповідають різним s. Нахил цих прямих поступово збільшується від нульового при $s = 0$ до максимального при $s{=}m$. Очевидно, що всі прямі, які відповідають $s \geq s_0$, де s_0 – складність істинної структури, проходитимуть через початок координат. На рис.2 пряма, що відповідає істинній структурі $s_0 = 3$, зображена потовщеною лінією, а пунктирною – пряма для $s = 4$. Зазначимо, що випадок $s = 0$ відповідає моделі з нульовими коефіцієнтами, тобто використанню вектора y в якості “моделі” $\hat{y}_{s|s=0} \equiv y$, тому на рис.1 всі графіки виходять з однієї точки $J(0 \mid \sigma^2 = 0) = J^b(0)$.

Графіки на рис.2 дають принципово нову інформацію: положення кожної прямої та точки їхнього перетину однозначно визначаються властивостями модельованого об’єкта, точніше, даними вибірки, і не залежать від характеристик шуму. При цьому огинаюча знизу цих графіків (позначена пунктирними лініями) є екстремаллю критерію $J(s)$: кожна точка на ній дорівнює мінімальному значенню критерію за відповідної дисперсії шуму. Отже, така огинаюча є лінією “перемикання” складності оптимальних структур: у разі збільшення дисперсії шуму (починаючи з $\sigma^2 = 0$) спочатку оптимальною буде істинна структура (тут $s_0 = 3$), після досягнення першої точки перемикання – структура на одиницю меншої складності ($s = 2$), і т.д., до $s = 0$.

Точки перемикання є нерухомими *критичними* точками, характерними для даної вибірки, тому природно ввести так звані ***критичні дисперсії*** [4] $\sigma^2_{kp}(s, s+1)$ для позначення координат цих точок. Як видно з рис.2, вони визначаються як розв’язок відносно σ^2 рівнянь

$$\sigma^2_{kp}(s,s+1):\ \ J(\sigma^2 \mid s) = J(\sigma^2 \mid s+1),\ \ s = 1, 2, \ldots, m-1\,, \tag{9}$$

які відповідають необхідній умові екстремуму функції дискретного аргумента. При цьому зрозуміло, що за заданої дисперсії шуму оптимальною буде структура s, для якої виконується нерівність

$$\sigma^2_{kp}(s-1, s) > \sigma^2 \geq \sigma^2_{kp}(s, s+1), \tag{10}$$

і навпаки – структура складності s буде оптимальною для σ^2, яка задовольнятиме цю нерівність.

Дисперсія помилки прогнозування $J_F(s)$ також лінійно залежить від σ^2, тому рис.2 актуальний і в цьому випадку, тобто можна означити критичні дисперсії $\sigma^2_{Fkp}(s, s+1)$ з умови, подібної до (9).

Сукупність процедур з обчислення критичних дисперсій, аналізу їхніх властивостей та інтерпретації отриманих результатів у задачі оптимізації структур моделей називається *методом критичних дисперсій*.

Наведемо загальну характеристику зв'язку основних властивостей величин критичних дисперсій з поведінкою оптимальних значень s^o, s_B^o , тобто з екстремальними властивостями критеріїв $J(s)$, $J_F(s)$: а) структурна складова критерію строго *монотонно* спадає, якщо для всіх $s = \overline{1, m-1}$ критичні дисперсії строго *додатні*; б) критерій має *єдиний мінімум* тільки тоді, коли всі значення критичних дисперсій строго монотонно зменшуються:

$$\sigma_{kp}^2(0,1) > \sigma_{kp}^2(1,2) > \ldots > \sigma_{kp}^2(m-1,m) \; ; \qquad (11)$$

в) при цьому *глобальний* мінімум критерію визначається умовою (10); г) якщо (11) не виконується, то (10) є умовою *локального* мінімуму, і для оптимальна складність відповідає мінімальному значенню дисперсії помилки моделі серед усіх точок локального мінімуму.

Аналіз реальних критеріїв вибору моделей

Ефективність заданого критерію $CR(s)$ аналізується з точки зору того, як відповідна йому оптимальна складність структури моделі

$$s_{CR}^* = arg \min_{s=\overline{1,m}} \mathrm{E}[CR(s)] \; , \qquad (12)$$

оцінює невідоме теоретичне значення s^o.

*Означення***.** Критерій $CR(s)$ називається *оптимальним* для оцінювання моделей з мінімальною дисперсією помилки, якщо $s_{CR}^* = s^o$ (незміщене оцінювання s^o), *адекватним* у разі $s_{CR}^* \le s^o$ (завадостійке оцінювання s^o) і *неадекватним,* якщо $s_{CR}^* > s^o$.

Тоді ефективність заданого критерію доцільно аналізувати так: знайти його математичне сподівання; перевірити компромісний характер структурної та шумової його складових; визначити умови оптимальності чи адекватності критерію. Останнє завдання вирішується формалізовано за допомогою методу критичних дисперсій.

Величина критичної дисперсії $\sigma_{CR}^2(s, s+1)$ як теоретична міра розрізнюваності моделей складності s та s+1 за критерієм $CR(s)$ дорівнює такому значенню σ^2, яке є розв'язком рівняння

$$\mathrm{E}[CR(s,\sigma^2)] = \mathrm{E}[CR(s+1,\sigma^2)] \; . \qquad (13)$$

Це дозволяє аналізувати ефективність заданого $CR(s)$, ґрунтуючись на такому результаті [5].

Твердження. Критерій $CR(s)$ є оптимальним чи адекватним, якщо виконуються умови $\sigma_{CR}^2(s, s+1) = \sigma_{kp}^2(s, s+1)$ або відповідно $\sigma_{CR}^2(s, s+1) \le \sigma_{kp}^2(s, s+1)$; інакше він є неадекватним.

Це твердження дає необхідні й достатні умови оптимальності чи адекватності критеріїв і для їхнього порівняння.

Актуальні задачі теорії структурної ідентифікації

Метод критичних дисперсій може допомогти у розв'язанні низки актуальних задач теорії: порівняльний аналіз критеріїв, планування експерименту, асимптотичний аналіз тощо. Розглянемо коротко деякі з отриманих результатів у цих задачах.

Порівняльний аналіз критеріїв. Приклади застосування методу критичних дисперсій для аналітичного порівняння кількох відомих критеріїв – Маллоуза C_p , Акаіке *FPE* та критерію регулярності $AR(s)$, який застосовується в МГУА, можна знайти у [5]. Показано, що у разі короткої вибірки критерій *FPE* є неадекватним, $AR(s)$ – адекватним, а C_p може бути адекватним, оптимальним і неадекватним у залежності від точності оцінки дисперсії шуму.

Задача планування експерименту. На основі аналізу критичних дисперсій встановлено [4], що ідентифікація за умовою мінімуму критерію $J_F(s)$ у разі заданого $X_F \neq X$ передбачає планування експерименту для досягнення пропорційності інформаційних матриць:

$$\rho^2 X^T X = X_F^T X_F, \quad \rho^2 \neq 0 \ . \tag{14}$$

Якщо матриці X та X_F такі, що кожна з них має взаємно ортогональні стовпці, то незалежно від їхнього співвідношення оптимальною для задач відновлення і прогнозування є одна й та ж модель.

Асимптотичний аналіз. Доведено, що необхідною й достатньою умовою асимптотичної незміщеності структури, оптимальної за заданим критерієм, є розбіжність критичної дисперсії цього критерію:

$$\sigma_{CR}^2(n,s,s+1) \xrightarrow[n]{} \infty. \tag{15}$$

За виконання традиційної для асимптотичного аналізу умови сильної регулярності регресорів критичні дисперсії як для ідеального критерію (3), так і для згаданих вище критеріїв, є розбіжними, тобто ці критерії є асимптотично оптимальними.

Аналіз умов оптимальності істинної структури. Показано, що умови оптимальності точної структури моделі у разі оцінювання параметрів за короткою вибіркою даних визначаються нерівністю:

$$\sigma_{кр}^2(s_0-1,s_0) \equiv \vartheta_{s_0}^2 \beta_{s_0} \geq \sigma^2 \ , \tag{16}$$

яка характеризує взаємозв'язок між точним параметром об'єкта, довжиною вибірки, планом експерименту та дисперсією (рівнем) шуму.

Висновок

Метод критичних дисперсій є ефективним апаратом теорії структурної ідентифікації моделей складних об'єктів в умовах неповноти інформації, причому як за обмеженої вибірки, так і в асимптотиці.

Список джерел

1. Современные методы идентификации систем. - М.: Мир, 1983. - 421 с.
2. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка, 1985. – 216 с.
3. Mallows C.L. Some comments on C_p // Technometrics. - 1973. - V.15. - P.661-667.
4. Степашко В.С. Структурная идентификация прогнозирующих моделей в условиях планируемого эксперимента // Автоматика. – 1992.– №1. – С.26-35.
5. Степашко В.С. Анализ эффективности критериев структурной идентификации прогнозирующих моделей // Проблемы управления и информатики. – 1994. – № 3-4. – С.13-22.