УДК 004.772:681.513

**Ткачов В. М.** *Харківський національний економічний університет ім. С. Кузнеця*

# МОДИФІКАЦІЯ МОДЕЛІ ПІДСИСТЕМИ ПРОМІЖНОГО ЗБЕРІГАННЯ В УМОВАХ ВІДСУТНОСТІ АПРІОРНОЇ ІНФОРМАЦІЇ ПРО ХАРАКТЕР BigData

Розглянуті питання функціонування інфокомунікаційних систем з огляду на проблему передачі великих масивів даних. Проаналізована базова модель підсистеми проміжного зберігання даних та вказані шляхи її модифікації для уніфікації при вирішенні задачі створення модульних систем передачі даних. Актуалізовано застосування методики раннього виявлення розладнань на основі гарантованого еліпсоїдального оцінювання до моделі підсистеми проміжного зберігання. Виконана модифікація моделі підсистеми проміжного зберігання.

***Ключові слова***: проміжне зберігання, апріорна інформація, BigData, інфокомунікаційна система

**Tkachov V. M. Modification of intermediate storage subsystem models in the absence of a priori information on the nature of BigData.** The questions of functioning information and communication systems due to the problem of transferring large amounts of data. Analyzed the basic model subsystems intermediate storage and the ways of its modifications for unification with the task of creating modular data transmission systems. Modified application techniques for early detection of disordered guaranteed ellipsoidal based assessment model to intermediate storage subsystem. Made modification of the model intermediate storage subsystem.

***Keywords***: intermediate storage priori information, BigData, information and communication system

**1. Постановка задачі.** Імпульс у розвитку фундаментальних наук на сьогоднішній день внесли сучасні комп’ютерні технології. Генерація великих обсягів інформації (BigData) під час досліджень дає можливість здобувати нові знання та виявляти нові закономірності [1].

Розвиток напрямів з передачі, зберігання та обробки даних типу BigData є однією з ключових задач розвитку IT-індустрії. Варто зазначити, що сучасні системи обробки даних (СОД), створені для проведення високошвидкісних обчислень, характеризуються розподіленістю та ступенем паралелізму. При цьому важливим класом задач мегасистем отримання нових знань є доставка даних до СОД. Однак в більшості організацій світу, що є генераторами даних типу BigData, передача даних здійснюються із застосуванням власних каналів зв'язку з низькою продуктивністю, що ускладнює проведення передачі і своєчасної координованої обробки цих даних при вирішенні розподілених задач.

Цій проблемі присвячено ряд наукових публікацій. Найбільш актуальною методикою для вирішення цієї проблеми є запровадження систем проміжного зберігання, що добре описано в роботах[2-4]. Суть полягає у використанні додаткових підсистем проміжного зберігання даних в низькошвидкісних ділянках інфокомуніційних мереж. При такій постановці завдання мінімізації часу передачі даних параметрами управління передачею даних є розміщення проміжних серверів зберігання даних та управління їх роботою.

При цьому середовищем для передачі даних є уже існуючі інфокомунікаційні мережі різної топології та архітектури, віртуалізовані рішення тощо. Теоретичне узагальнення та ряд прикладних експериментів дали змогу виявити підзадачі, вирішення яких допомагають провести заходи з оптимізації методу та його уніфікації [5].

Задачу поточної побудови моделі підсистеми проміжного зберігання (ідентифікації)як об’єкта керування можна представити у вигляді лінійного рівняння

$$y(t) = \theta^T x(t) + w(t),$$

де збурення $w(t)$ має статистичний характер і частіше за все це циркуляція потоків даних з нульовим математичним сподіванням і скінченною дисперсією, на кшталт білого гаусівського шуму в моделях керування об’єктами. Таке припущення створює клас алгоритмів поточної ідентифікації [9]. Інші припущення про статичний характер циркуляцій потоків даних в мережі вимагають використання критеріїв оцінювання, відмінних від суми квадратів похибок. Це дає можливість застосовувати алгоритми ідентифікації, засновані на методах баєсівського оцінювання, найменших модулів тощо [10, 11].

Однак, в більшості праць з проблеми ідентифікації в умовах імпульсивних циркуляцій потоків даних припускається, що ці імпульси є випадковими у відношенні до підсистеми, а отже і оцінки невідомого вектору $\theta$ можуть бути знайдені лише у формі моментів деякого багатовимірного розподілу.

Останні роки у більшості праць не описується характер циркуляцій потоків даних (більш того, циркуляцій потоків даних можуть мати регулярний детермінований характер), крім їх належності деякому обмеженому інтервалу.

Одне з важливих питань, яке потребує вирішення, полягає у створенні підґрунтя для модифікації підсистеми проміжного зберігання в умовах відсутності апріорної інформації про характер даних типу BigData. У даній роботі пропонуються теоретичні варіанти вирішення задачі шляхом математичного моделювання та створення модифікованої моделі підсистеми проміжного зберігання.

**2. Узагальнення моделі підсистеми проміжного зберігання за принципом методу Лагранжа.** Розглянемо статичну модель, призначену для модуля керування ємністю підсистеми збереження даних (рис. 1), що включає $n(n>1)$ видів даних за пріоритетом передачі, що зберігаються в підсистемі збереження даних з обмеженою ємністю. Дану умову, яка визначає взаємозв'язок між різними видами даних, включено в модель як обмеження.

Позначимо: $A$ – максимально допустима ємність підсистеми збереження даних для $n$ видів даних; $a_i$ – ємність, яка необхідна для збереження однієї порції даних $i$-го виду; $y_i$ – запит на передачу деякої величини порції $i$-го виду даних. У такому разі обмеження на ємність підсистеми збереження даних складає:

$$\sum_{i=1}^{n} a_i y_i \le A .$$


Рис. 1. Структура підсистеми збереження даних

У базовій моделі [2] декларується, що об'єм даних кожного виду надходить протягом часу, який визначається для кожної порції в залежності від завантаження загальної пропускної здатності каналу передачі даних, згідно темпорального критерію [6]. Враховуючи це, припустимо, що час для кожного мінімального інтервалу є константою.

Позначимо: $K_i$ – витрати на формування запиту на зберігання порцій даних; $\beta$ – інтенсивність передачі порцій даних від одного джерела (в залежності від типу даних); $h_i$ – витрати (зайнятість деякого об'єму) на збереження порції даних за одиницю часу для кожного $i$-го виду даних.

Таким чином, задача узагальнення моделі підсистеми проміжного зберіганняна буває вигляду:

$$f(y_1, y_2, ..., y_n) = \sum_{i=1}^{n} \frac{K_i \beta_i}{y_i} + h_i \left( \frac{y_i}{2} \right) \to \min \tag{1}$$

$$\text{при } \sum_{i=1}^{n} a_i y_i \le A,\ y_i > 0, i = \overline{1, n}. \tag{2}$$

Загальний розв'язок даної задачі можна знайти методом множників Лагранжа [7]. Проте перш ніж застосовувати метод, потрібно з'ясувати, чи задовольняє обмеження на ємність підсистеми збереження даних розв'язок задачі пошуку безумовного мінімуму функції (1), яка відповідає задачам (1), (2).

Градієнт функції цілі (1) має вигляд:

$$\nabla f = \left( -\frac{K_1 \beta_1}{y_1^2} + \frac{h_1}{2}, -\frac{K_2 \beta_2}{y_2^2} + \frac{h_2}{2}, ..., -\frac{K_n \beta_n}{y_n^2} + \frac{h_n}{2} \right).$$

Стаціонарна точка $y^* = (y_1^*, y_2^*, ..., y_n^*)$ функції цілі (1), в якій градієнт функції дорівнює нулеві, має координати:

$$y_i^* = \sqrt{\frac{2K_i \beta_i}{h_i}}. \tag{3}$$

Матриця Гессе функції цілі (1) є діагональною вигляду:

$$\begin{bmatrix} 2\frac{K_1\beta_1}{y_1^3} & 0 & ... & 0 \\ 0 & 2\frac{K_2\beta_2}{y_2^3} & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & 2\frac{K_n\beta_n}{y_n^3} \end{bmatrix}.$$

Всі діагональні елементи більші від нуля. Легко помітити, що за критерієм Сільвестра ця матриця є додатною визначеною [7] і вектор (3) визначає точку глобального мінімуму задачі $I(x^*) = \{ i \mid h_i(x^*) = 0, 1 \le i \le m \}$.

Якщо в точці (3) обмеження (2) виконується, то воно надлишкове і його можна не враховувати.

Обмеження діє, якщо воно не виконується для значень $y_i^*$. Тут потрібно знайти нове оптимальне значення $y_i^*$, яке задовольняє обмеження на ємність підсистеми збереження даних, що накладається у вигляді рівності.

У даному випадку така модифікація задачі можлива, оскільки функція цілі опукла і задача має одне лінійне обмеження. Така редукція задачі може виявитися некоректною за інших обмежень, або якщо їх більше, ніж одне. У такому випадку, для вирішення задачі узагальнення необхідно застосовувати інші методики побудови моделі.

Отже, для розв'язування даної задачі можна застосувати викладену вище теорію. Побудуємо функцію Лагранжа вигляду:

$$L(\lambda, y_1, y_2, ..., y_n) = f(\lambda, y_1, y_2, ..., y_n) - \lambda \left( \sum_{i=1}^{n} a_i y_i - A \right) = \sum_{i=1}^{n} \left( \frac{K_i \beta_i}{y_i} + h_i \frac{y_i}{2} \right) - \lambda \left( \sum_{i=1}^{n} a_i y_i - A \right),$$

де $\lambda (< 0)$ – множник Лагранжа.

Оптимальні значення $y_i^*$ і $\lambda$ можна знайти, прирівнявши до нуля відповідні часткові похідні функції Лагранжа, що дає систему нелінійних рівнянь вигляду:

$$\frac{\partial L}{\partial y_i} = -\frac{K_i \beta_i}{y_i^2} + \frac{h_i}{2} - \lambda a_i = 0, i = \overline{1,n}, \quad \frac{\partial L}{\partial \lambda} = -\sum_{i=1}^{n} a_i y_i + A = 0 ,$$

Розв'язування побудованої системи нелінійних рівнянь може здійснюватися одним із варіацій чисельних методів. Таким чином, вигляд узагальненої моделі підсистеми проміжного зберігання за принципом методу Лагранжа може приймати рішення, які адаптуються для поставленої задачі даної роботи – модифікація моделі підсистеми проміжного зберігання в умовах відсутності апріорної інформації про характер BigData.

**4. Застосування методики раннього виявлення розладнань на основі гарантованого еліпсоїдального оцінювання до моделі підсистеми проміжного зберігання**. Основні принципи методики раннього виявлення розладнань на основі гарантованого еліпсоїдального оцінювання детально розглядаються у праці [8].

Розглянемо:

$$y(t) = \theta^T x(t) + w(t), \; | w(t) | \le r(t), t = 1, 2, ..., n, ..., \tag{4}$$

де обмеження $r(t)$ може вибиратися достатньо великим і зменшуватися в процесі ідентифікації (задати межі циркуляцій потоків даних у більшості практичних ситуацій – не складно). Переписуючи (4) у вигляді:

$$y(t) - r(t) \le \theta^T x(t) \le y(t) + r(t) , \tag{5}$$

можна помітити, що ці нерівності задають пару гіперповерхонь у просторі $\theta$, між якими знаходиться невідомий вектор параметрів. Послідовність спостережень за станом інфокомунікаційної мережі $y(1), y(2), ..., y(n)$, породжує $n$ пар гіперповерхонь, що створюють у просторі $\theta$ висічену область $D_n$. Це – область оцінок $\hat{\theta}(t)$. Сукупність точок цієї області – рівноправні (5). Тобто, результатом вирішення задачі оцінювання буде інтервальна точкова оцінка [8].

Очевидним рішенням задачі є знаходження розв'язку системи $n$ лінійних нерівностей типу [57]. Але, виходячи з того, що кількість вершин політопа $D_n$ зростає значно швидше, ніж $t = 1, 2, ..., n, ...$ з точки зору організації обчислень – цей підхід буде мати низькі показники ефективності. Пропонується альтернативний підхід [8], який полягає в апроксимації політопа $D_n$, що отримано в $t$-й поточний момент часу, еліпсоїдом

$$E_t : \left(\theta - \hat{\theta}(t)\right)^T p^{-1}(t)\left(\theta - \hat{\theta}(t)\right) \le 1, \tag{6}$$

центр якого $\hat{\theta}(t)$ та семерична додатна визначена матриця $P(t)$ уточняються так, щоб $E_t$ був в області $D_n$. Однак $\hat{\theta}(t)$ і $P(t)$ містять лише $n_\theta + \left(n_\theta + 1\right) n_\theta / 2$ параметрів, які можна налаштовувати. Тому цей алгоритм має істотні переваги.

Ідея підходу (підхід Ф. Швепе [12]), полягає в наступному: еліпсоїд $E_t$ включає всі оцінки, що належать перетину $E_{t-1}$ з областю $F_t$; ця область лежить між двома гіперповерхнями останнього спостереження (4) (рис 2).

Оскільки перетин $E_{t-1}$ і $F_t$ не є еліпсоїдом, необхідно так побудувати $\hat{\theta}(t)$ і $P(t)$, щоб $E_t$ максимально точно його апроксимував.

Об'єднавши(4) і (6), невідомі параметри можна представити у вигляді системи нерівностей:

$$\begin{cases} \left(\theta - \hat{\theta}(t)\right)^T p^{-1}(t)\left(\theta - \hat{\theta}(t)\right) \le 1 \\ r^{-2}(t)\left(y(t) - \theta^T x(t)\right)^2 \le 1 \end{cases} .$$

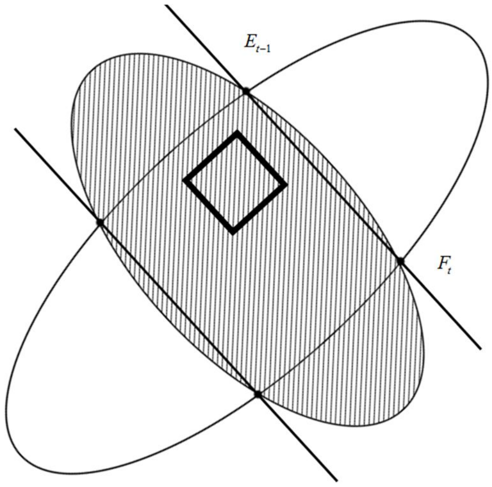


Рис. 2. Апроксимація перетину еліпсоїда парою поверхонь

Якщо величина $p(t)$ приймає будь-яке невід'ємне значення, то:

$$\left(\theta-\hat{\theta}(t)\right)^T p^{-1}(t)\left(\theta-\hat{\theta}(t)\right)+p(t)r^{-2}(t)\left(y(t)-\theta^T x(t)\right)^2 \le 1+p(t) \qquad (7)$$

Для спрощення виразу проведемо деякі очевидні перетворення квадратичної форми в лівій частині (7). Вводячи похибку оцінювання, яка може виникати з огляду на бітові помилки, що можуть виникати при негативних відповідях підсистеми щодо цілісності даних (рис. 1) $\tilde{\theta}(t)=\theta-\hat{\theta}(t)$:

$$\theta^T x(t)=\hat{\theta}(t)(t-1)x(t)+\left(\theta-\hat{\theta}(t-1)\right)^T x(t)=\hat{\theta}^T(t-1)x(t)+\tilde{\theta}^T(t-1)x(t),$$

$$\tilde{\theta}^T(t-1)P^{-1}(t-1)\tilde{\theta}(t-1)+p(t)r^{-2}(t)\left(y(t)-\hat{\theta}^T(t-1)x(t)-\tilde{\theta}^T(t-1)x(t)\right)\le 1+p(t),$$

$$\tilde{\theta}^T(t-1)P^{-1}(t-1)\tilde{\theta}(t-1)+p(t)r^{-2}(t)\left(v(t)-\tilde{\theta}^T(t-1)x(t)\right)^2\le 1+p(t),$$

де $v(t)=y(t)-\hat{\theta}^T(t-1)x(t)$ – похибка ідентифікації[8];

$$\tilde{\theta}^T(t-1)P^{-1}(t-1)\tilde{\theta}(t-1)+p(t)r^{-2}(t)v^2(t)-2p(t)r^{-2}(t)v(t)\tilde{\theta}^T(t-1)x(t)+$$
$$+p(t)r^{-2}(t)\tilde{\theta}^T(t-1)x(t)x(t)^T(t)\tilde{\theta}(t-1)\le 1+p(t),$$
$$\tilde{\theta}^T(t-1)\left(P^{-1}(t-1)+p(t)r^{-2}(t)x(t)x^T(t)\right)\tilde{\theta}(t-1)+$$
$$+p(t)r^{-2}(t)v^2(t)-2p(t)r^{-2}(t)v(t)\tilde{\theta}^T(t-1)x(t)\le 1+p(t).$$

Позначимо $\tilde{P}^{-1}(t-1)=P^{-1}(t-1)+p(t)r^{-2}(t)x(t)x^T(t)$:

$$\tilde{\theta}^T(t-1)\tilde{P}^{-1}(t-1)\tilde{\theta}(t-1)-2p(t)r^{-2}(t)v(t)\tilde{\theta}^T(t-1)x(t)+p(t)r^{-2}(t)v^2(t)\le 1+p(t). \qquad (8)$$

Доповнимо (8) до повного квадрата:

$$\tilde{\theta}^T(t-1)\tilde{P}^{-1}(t-1)\tilde{\theta}(t-1)-2p(t)r^{-2}(t)v(t)\tilde{\theta}^T(t-1)x(t)+p^2(t)r^{-4}(t)v^2(t)\times$$
$$\times x^T(t)\tilde{P}(t-1)x(t)-p^2(t)r^{-4}(t)v^2(t)x^T(t)\tilde{P}(t-1)x(t)+p(t)r^{-2}(t)v^2(t)\le 1+p(t),$$
$$\left(\tilde{\theta}^T(t-1)-p(t)r^{-2}(t)v(t)\tilde{P}(t-1)x(t)\right)^T\tilde{P}^{-1}(t-1)\left(\tilde{\theta}(t-1)-p(t)r^{-2}(t)v(t)\tilde{P}(t-1)x(t)\right)+$$
$$+p(t)r^{-2}(t)v^2(t)-p^2(t)r^{-4}(t)v^2(t)x^T(t)\tilde{P}(t-1)x(t)\le 1+p(t),$$

після чого, вводячи позначення $\tilde{\theta}(t)=\tilde{\theta}(t-1)-p(t)r^{-2}(t)v(t)\tilde{P}(t-1)x(t)$, отримуємо:

$$\tilde{\theta}^T(t)\tilde{P}^{-1}(t-1)\tilde{\theta}(t)+p(t)r^{-2}(t)v^2(t)-p^2(t)r^{-4}(t)v^2(t)x^T(t)\tilde{P}(t-1)x(t)\le 1+p(t) \qquad (9)$$

і

$$\hat{\theta}(t)=\hat{\theta}(t-1)+p(t)r^{-2}(t)v(t)\tilde{P}(t-1)v(t)x(t), \qquad (10)$$

або

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)\left(y(t) - \hat{\theta}^T(t-1)x(t)\right)x(t),$$

що співпадає із загальною схемою рекурентної побудови моделі підсистеми проміжного зберігання(8), (9) [8, 9], хоча матричний коефіцієнт підсилення алгоритму $\gamma(t) = p(t)r^{-2}(t)v(t)\tilde{P}(t-1)$ – не визначено, оскільки конкретне значення $p(t)$ – теж невідоме.

Повертаючись до введеного виразу $\tilde{P}(t-1) = P^{-1}(t-1) + p(t)r^{-2}(t)x(t)x^T(t)$, який з використанням леми обернення матриць Шермана-Морісона [13] може бути записаний у вигляді

$$\tilde{P}(t-1) = P(t-1) - \frac{p(t)r^{-2}(t)P(t-1)x(t)x^T(t)P(t-1)}{1 + p(t)r^{-2}(t)x^T(t)P(t-1)x(t)}, \qquad (11)$$

перевіримо нерівність (9)

$$\begin{aligned}
&\tilde{\theta}^T(t)\tilde{P}^{-1}(t-1)\tilde{\theta}(t) + p(t)r^{-2}(t)v^2(t) - p^2(t)r^{-4}(t)x^T(t)\tilde{P}(t-1)x(t) = \\
&= \tilde{\theta}^T(t)\tilde{P}^{-1}(t-1)\tilde{\theta}(t) + p(t)r^{-2}(t)v^2(t)\left(1 - r^{-2}(t)p(t)x^T(t)\tilde{P}(t-1)x(t)\right) = \\
&= \tilde{\theta}^T(t)\tilde{P}^{-1}(t-1)\tilde{\theta}(t) + p(t)r^{-2}(t)v^2(t)\Bigg(1 - p(t)r^{-2}(t)x^T(t) \times \\
&\times \frac{P(t-1) - p(t)r^{-2}(t)P(t-1)x^T(t)P(t-1)x(t) - p(t)r^{-2}(t)P(t-1)x(t)x^T(t)P(t-1)}{1 + p(t)r^{-2}(t)x^T(t)P(t-1)x(t)}x(t)\Bigg) = \\
&= \tilde{\theta}^T(t)\tilde{P}^{-1}(t-1)\tilde{\theta}(t) + \frac{p(t)r^{-2}(t)v^2(t)}{1 + p(t)r^{-2}(t)x^T(t)P(t-1)x(t)},
\end{aligned}$$

або

$$\left(\theta - \hat{\theta}(t)\right)^T \tilde{P}^{-1}(t-1)\left(\theta - \hat{\theta}(t)\right) \le 1 + p(t) - \frac{p(t)r^{-2}(t)v^2(t)}{1 + p(t)r^{-2}(t)x^T(t)P(t-1)x(t)}.$$

Вводячи позначення

$$P(t) = \tilde{P}(t-1)\left(1 + p(t) - \frac{p(t)v^2(t)}{r^2(t) + p(t)x^T(t)P(t-1)x(t)}\right)$$

та використовуючи співвідношення (10) та (11), отримаємо вирази, що задають алгоритм рекурентної ідентифікації при обмеженому характеру циркуляцій потоків даних:

$$\begin{cases}
\hat{\theta}(t) = \hat{\theta}(t-1) + p(t)r^{-2}(t)\tilde{P}(t-1)\left(y(t) - \hat{\theta}^T(t-1)x(t)\right)x(t), \\
\tilde{P}(t) = P(t-1) - \dfrac{p(t)r^{-2}(t)P(t-1)x(t)x^T(t)P(t-1)}{1 + p(t)r^{-2}(t)x^T(t)P(t-1)x(t)}, \\
P(t) = \tilde{P}(t-1)\left(1 + p(t) - \dfrac{p(t)\left(y(t) - \hat{\theta}^T(t-1)x(t)\right)^2}{r^2(t) + p(t)x^T(t)P(t-1)x(t)}\right).
\end{cases} \qquad (12)$$

Алгоритм (12) є різновидом зваженого методу найменших квадратів і структурно є близьким до алгоритму Фогеля-Хуанга [14]. У ньому є невизначений параметр $p(t)$, який обирається з огляду на те, щооб'єм еліпсоїда $E_t$, що описує перетин $E_{t-1}$ і $F_t$, має бути найменшим. Об'єм еліпсоїда (6) $V_t$ пропорційний значенню $\sqrt{\det P(t)}$ :

$$V_t \sim \sqrt{\det P(t)} = \prod_{i=1}^{n_\theta} \lambda_i^{0,5}$$

де $\lambda_i (i = 1, 2, ..., n_\theta)$ – власні числа матриці $\mathrm{P(t)}$.

Задача пошуку значення $p(t)$ розглядається як задача мінімізації $\det P(t)$ по $p(t)$ (задача одновимірного пошуку). Критерій, пов'язаний з $\det P(t)$ є D-критерієм.

Виразимо $\det P(t)$ як функцію $r(t)$. Для цього проведемо наступні матричні перетворення:

$$\tilde{P}(t-1) = P(t-1) - \frac{p(t)r^{-2}(t)P(t-1)x(t)x^T(t)P(t-1)}{1+p(t)r^{-2}(t)x^T(t)P(t-1)x(t)} = \left(1 - \frac{p(t)P(t-1)x(t)x^T(t)}{r^2 + p(t)x^T(t)P(t-1)x(t)}\right)P(t-1), \quad (13)$$

$$\det \tilde{P}(t-1) = \det\left(1 - \frac{p(t)P(t-1)x(t)x^T(t)}{r^2 + p(t)x^T(t)P(t-1)x(t)}\right)\det P(t-1).$$

Позначаючи

$$\begin{cases} p(t)P(t-1)x(t) = b(t) \\ -\dfrac{x(t)}{r^2(t) + p(t)x^T(t)P(t-1)x(t)} = c(t) \\ x^T(t)P(t-1)x(t) = g(t) \end{cases}$$

з урахуванням матричної тотожності $\det\left(1 + b(t)c^T(t)\right) = 1 + c^T(t)b(t)$ одержуємо:

$$\det \tilde{P}(t) = \left(1 - \frac{p(t)g(t)}{r^2(t) + p(t)g(t)}\right)\det P(t-1),$$

$$\det P(t) = \left(1 + p(t) - \frac{p(t)v^2(t)}{r^2(t) + p(t)g(t)}\right)^{n_\theta}\left(1 - \frac{p(t)g(t)}{r^2(t) + p(t)g(t)}\right)\det P(t-1). \quad (14)$$

Оскільки рівняння

$$\frac{\partial \det P(t)}{\partial p(t)} = 0 \quad (15)$$

не має аналітичного розв’язку, застосовуємо процедуру одновимірного пошуку мінімума (14), або процедуру пошуку дійсних додатних коренів (15).

Процедура (12) реалізується за допомогою послідовності кроків:

1. Задається $\hat{\theta}(0) = 0, P(0) = \gamma, \gamma \geq 1$.

2. Знаходяться $g(t) = x^T(t)P(t-1)x(t)$ і $v(t) = y(t) - \hat{\theta}^T(t-1)x(t)$.

3. Вирішується задачу мінімізації (14). Якщо додатне значення $p(t)$ не знайдене, це означає, що $E_{t-1}$ і $F_t$ – не перетинаються,а $t$ реєструється як момент розладнання в роботі підсистеми. Щоб процедура працювала далі, після фіксації розладнання адають як $p(t) = 0$, тобто спостереження $y(t)$ ігнорується. Якщо знайдено декілька додатних коренів рівняння (15), серед них обирається той, який забезпечує найменше значення величини

$$\left(1 + p(t) - \frac{p(t)v^2(t)}{r^2(t) + p(t)g(t)}\right)^{n_\theta}\left(1 - \frac{p(t)g(t)}{r^2(t) + p(t)g(t)}\right).$$

4. Знаходиться $\tilde{P}(t-1)$ з виразу (13).

5. Знаходиться $\hat{\theta}(t) = \hat{\theta}(t-1) + p(t)r^{-2}(t)\tilde{P}(t-1)v(t)x(t)$.

6. Знаходиться $P(t) = \tilde{P}(t-1)\left(1 + p(t) - \dfrac{p(t)v^2(t)}{r^2(t) + p(t)g(t)}\right)$.

7. У разі відсутності рішення – крок 2. Кількість ітерацій проходження задається на початковому етапі роботи алгоритму.

З метою спрощення чисельної реалізації алгоритму (12) перейдемо до рівняння [8]:

$$\hat{\theta}(t) = \beta(t)\left( P(t-1) - \frac{P(t-1)x(t)x^T(t)P(t-1)}{\alpha(t) + x^T(t)P(t-1)x(t)} \right),$$

де $\alpha(t) = p^{-1}(t)r^2(t)$, $\beta(t) = 1 + \frac{r^2(t)}{\alpha(t)} - \frac{v^2(t)}{\alpha(t) + x^T(t)P(t-1)x(t)}$.

В результаті отримуємо алгоритм у вигляді

$$\begin{cases} \hat{\theta}(t) = \hat{\theta}(t-1) + \dfrac{P(t-1)\left( y(t) - \hat{\theta}^T(t-1)x(t) \right)x(t)}{\alpha(t) + x^T(t)P(t-1)x(t)} \\ P(t) = \beta(t)\left( P(t-1) - \dfrac{P(t-1)x(t)x^T(t)P(t-1)}{\alpha(t) + x^T(t)P(t-1)x(t)} \right) \end{cases},$$

який структурно повністю співпадає з експоненційно зваженим рекурентним методом найменших квадратів, але відрізняється від нього своїми властивостями. На кожному кроці необхідно вирішувати задачу мінімізації функції:

$$\det P(t) = \left( 1 + \frac{r^2(t)}{\alpha(t)} - \frac{v^2(t)}{\alpha(t) + g(t)} \right)^{n_\theta} \left( 1 - \frac{g(t)}{\alpha(t) + g(t)} \right) \det P(t-1).$$

Необхідність вирішення задачі мінімізації функції (14) суттєво ускладнює використання розглянутого алгоритму в процедурах ранньої діагностики циркуляцій потоків даних. В таких випадках є доцільним застосування методики виявлення циркуляцій потоків даних за допомогою спрощеного алгоритму еліпсоїдального оцінювання [8]. У такому разі, узагальнена форма алгоритму еліпсоїдального оцінювання набуває вигляду:

$$\begin{cases} \hat{\theta}(t) = \hat{\theta}(t-1) + \alpha(t)\delta(t)v(t)P(t)x(t), \\ P(t) = \dfrac{\alpha(t)}{\alpha(t-1)}\left( P(t-1) - \delta(t)\dfrac{P(t-1)x(t)x^T(t)P(t-1)}{\alpha(t-1) + \delta(t)x^T(t)P(t-1)x(t)} \right), \\ \alpha(t) = \alpha(t-1) + \delta(t)r^2(t) - \dfrac{\alpha(t-1)\delta(t)v^2(t)}{\alpha(t-1) + \delta(t)x^T(t)P(t-1)x(t)}, \\ 0 < \delta(t) \le \dfrac{\dfrac{v^2(t)}{r^2(t)} - 1}{x^T(t)P(t-1)x(t)} \end{cases},$$

контроль за циркуляцією потоків даних за допомогою якого ведеться на основі послідовності кроків, наведеної вище.

**5. Висновки.** У роботі вирішене часткове питання щодо створення модифікації підсистеми проміжного зберігання в умовах відсутності апріорної інформації про характер даних типу BigData. Запропоновані варіанти вирішення задачі лежать у площині математичного моделювання на подвійній платформі: узагальнення моделі підсистеми проміжного зберігання за принципом методу Лагранжа та вирішення задачі рекурентного створення моделі підсистеми проміжного зберігання в умовах відсутності апріорної інформації про характер BigData (з огляду на циркуляції потоків даних, як дестабілізуючого фактору при передачі даних в інфокомунікаційній мережі), крім її обмеженості, виявляти розладнання в роботі підсистеми збереження даних у онлайн-режимі. Зроблені висновки про простоту у чисельній реалізації і можливу модель функціонування, за якої підсистема збереження даних, накопичуючи інформацію про збурення в мережі, у процесі роботі протягом деякого інтервалу часу, поступово приймає вигляд відомих процедур.

Рекомендується продовжити дослідження у площині практичного застосування отриманих теоретичних напрацювань з метою виявлення класів задач щодо побудови модулів на базі інтелектуального прийняття рішень щодо адаптації підсистеми збереження даних у разі виникнення збурень в інфокомунікаційній мережі при передачі даних типу BigData.

**Список використаної літератури**


1. Katina M. Big Data: New Opportunities and New Challenges / Katina Michael, Keith W. Miller // IEEE Computer Society. – 2013. – Vol. 46, № 6. – P. 22-24.

2. Саваневич В.Е. Метод передачи данных с промежуточным хранением / В.Е. Саваневич, В.Н. Ткачев // Системы обработки информации: сборник научных трудов. – 2014. – № 7 (123). – С. 99-105.

3. Ткачев В.Н. Использование метода передачи данных с промежуточным хра-нением при передаче результатов радиоастрономических наблюдений / В.Н. Ткачев, А.М. Резниченко // Наукові записки Українського науково-дослідного інституту зв'язку. – 2014. – № 4 (32). – С. 49-53.

4. Tkachov V.M. Method for transfer of data with intermediate storage / V.M. Tkachov, V.Ye. Savanevych / IEEE First International Scientific-Practical Conference «Problems of Infocommunications. Science and Technology» (PICS&T-2014), Kharkiv, October 14-17, 2014. – 2014. – P. 105-106.

5. Ткачев В.Н. Разработка метода передачи данных с промежуточным хранением / В.Н. Ткачев, В.Е. Саваневич / 5 Международный радиоэлектронный форум "Прикладная радиоэлектроника. Состояние и перспективы развития" (МРФ-2014), Харьков, 14-17 октября 2014 г. – Харьков, 2014. – Т.2. – С. 157-159.

6. Tkachov V.M. Automated Controllers Functioning Criteria in Content Distribu-tion Systems / V.M. Tkachov, V.E. Savanevych // Scholars Journal of Engineer-ing and Technology. – 2014. – 2(3A). – P. 348-351.

7. Петров Е.Г. Методи і засоби прийняття рішень у соціально-економічних системах / Е.Г. Петров, М.В. Новожилова, І.В. Гребеннік. – К.: Техніка, 2004. – 256 с.

8. Бодянський Є.В. Адаптивне виявлення розладнань в об'єктах керування за допомогою штучних нейронних мереж / Є.В. Бодянський, О.І. Міхальов, І.П. Плісс. – Дніпропетровськ: Системні технології, 2000. – 140 с.

9. Льюинг Л. Идентификация систем. Теория для пользователя / Л. Льюинг. – М.: Наука, 1991. – 432 с.

10. Ципкин Я.З. Основы информационной теории идентификации / Я.З. Ципкин. – М.: Наука, 1984. – 320 с.

11. Рубан А.И. Исследование адаптивных робастных алгоритмов идентификации параметров нестационарных объектов / А.И. Рубан, Р.Р. Ялатдинов. – Томск, 1985. – 20 с. – Деп. в ВИНИТИ 11.07.1985., № 5637-85.

12. Schweppe F.C. Uncertain Dynamic Systems / F.C. Schweppe. – Englewood Cliffs., NJ.: Prentice-Hall, Inc., 1973. – 533 p.

13 Малоземов В.Н. Формулы Фробениуса, Шермана–Моррисона и близкие вопросы / В.Н. Малоземов, М.Ф. Монако, А.В. Петров // Журнал вычислительной математики и математической физики. – 2002. – Т. 42, № 10. – С. 1459-1465.

14. Fogel E. On the value of information in system identification- bounded noise case / E. Fogel, Y.F. Huang. – Automatika. – 1987. – V. 18, № 2. – P. 229.

**References**


1. Katina M. Big Data: New Opportunities and New Challenges / Katina Michael, Keith W. Miller // IEEE Computer Society. – 2013. – Vol. 46, № 6. – P. 22-24.

2. Savanevych V.Ye. Method of data transfer with intermediate storage / V.Ye. Savanevych, V.M. Tkachov // Information Processing Systems. – 2014. – № 7 (123). – P. 99-105.

3. Tkachov V.M. Using the method of data transfer with intermediate storage in the transmission of radio astronomical observations / V.M. Tkachov, O.M. Reznichenko // Scientific notes of «Ukrainian Research Institute of Communications»: scientific and production collection. – 2014. – № 4 (32). – P. 49-53.

4. Tkachov V.M. Method for transfer of data with intermediate storage / V.M. Tkachov, V.Ye. Savanevych / IEEE First International Scientific-Practical Conference «Problems of Infocommunications. Science and Technology» (PICS&T-2014), Kharkiv, October 14-17, 2014. – 2014. – P. 105-106.

5. Tkachov V.M. Development of the method of data transfer with intermediate storage / V.M. Tkachov, V.Ye. Savanevych / 5 International Radioelectronic Forum "Applied Radioelectronics: Status and Prospects of Development" (MRF-2014), Kharkiv, 14-17 october 2014 г. – Kharkiv, 2014. – V.2. – P. 157-159.

6. Tkachov V.M. Automated Controllers Functioning Criteria in Content Distribu-tion Systems / V.M. Tkachov, V.E. Savanevych // Scholars Journal of Engineer-ing and Technology. – 2014. – 2(3A). – P. 348-351.

7. Petrov Ye.G. Methods and tools for decision making in socio-economic systems / Ye.G. Petrov, M.V. Novozhilova, I.V. Grebennik. – K.: Tekhnika, 2004. – 256 p.

8. Bodyanskiy Ye.V. Adaptive detection of objects in disordered control using artificial neural networks / Ye.V. Bodyanskiy, O.I. Mikhalyov, I.P. Pliss. – Dnepropetrovsk: System technologies, 2000. – 140 p.

9. Luing L. Identification of systems. Theory for the user / L. Luing. – M.: Science, 1991. – 432 p.

10. Cypkin Ya. Z. Fundamentals of the Information Theory of Identification / Ya.Z. Cypkin. – M.: Science, 1984. – 320 p.

11. Ruban A.I. Study of adaptive robust algorithms for identifying parameters of non-stationary objects / A.I. Ruban, R.R. Yalatdinov. – Tomsk, 1985. – 20 p. – Dep. in VINITI 11.07.1985., № 5637-85.

12. Schweppe F.C. Uncertain Dynamic Systems / F.C. Schweppe. – Englewood Cliffs., NJ.: Prentice-Hall, Inc., 1973. – 533 p.

13 Malozemov V.N. Formulas of Frobenius, Sherman-Morrison and related questions / V.N. Malozemov, M.F. Monaco, A.V. Petrov // The Journal of Computational Mathematics and Mathematical Physics. – 2002. – V. 42, № 10. – P. 1459-1465.

14. Fogel E. On the value of information in system identification- bounded noise case / E. Fogel, Y.F. Huang. – Automatika. – 1987. – V. 18, № 2. – P. 229.



***Відомості про автора***

**Ткачов Віталій Миколайович** – кандидат технічних наук, викладач кафедри Інформаційних систем, Харківський національний економічний університет ім. С. Кузнеця. Тел. +380 (93) 785 02 87. E-mail: vitalii@tkachov.com

***Author of the article***

**Tkachov Vitalii Mykolayovych** – candidate of sciences (technical), teacher of the information systems department, Simon Kuznets Kharkiv National University of Economics. Tel.: +380 (93) 785 02 87. E-mail: vitalii@tkachov.com



Дата надходження в редакцію: 17.04.2017 р.

Рецензент: доктор технічних наук, професор К. С. Козелкова

*Державний університет телекомунікацій, Київ*