

## МАТЕМАТИЧНА МОДЕЛЬ ОБРОБКИ ДАНИХ ЗА КРИТЕРІЄМ МІНІМУМУ ПРОТЯЖНОСТІ ДЛЯ СКЛАДНИХ УМОВ СПОСТЕРЕЖЕННЯ

*Анотація.* Досліджена ефективність застосування математичної моделі обробки даних, що побудована за критерієм мінімуму протяжності відхилю розв'язку, для складних умов спостереження. Зазначена модель уможливіє налаштування процесу обробки даних на поточне шумове оточення, яке полягає у пошуку значень трьох вільних параметрів функціоналу квазіпротяжності, що найкраще узгоджують відому модель з отриманими даними на інтервалі спостереження. Описано метод налаштування обробки даних для випадку, коли моделлю даних є константа, та подані результати чисельного моделювання й рекомендації з ефективного налаштування.

*Ключові слова:* математична модель, функціонал, налаштування.

**Вступ.** Складні умови спостереження, які виникають за наявності шуму невідомої статистичної природи, потужних завад та/або раптових змін значень робочих параметрів досліджуваного об'єкта, приводять до отримання суттєво спотворених й неповних даних. Такі умови значно утруднюють процеси побудови й верифікації його математичної моделі через ускладнення проблеми узгодження заданої моделі та отриманих даних. Останнє обумовлене тим, що за складних умов спостереження для ефективного узгодження моделі й даних необхідно, по-перше, визначити ті елементи даних, які відповідають заданій моделі з очікуваними значеннями її параметрів, та виключити з подальшого розгляду неправдоподібні значення або ж значно послабити їхній вплив і, по-друге, встановити наявність фрагментів даних, які відповідають шуканій моделі з іншими значеннями її параметрів. Якщо перша вимога може бути реалізована на основі робастного підходу [1], то друга вимога диктує пошук певної множини локальних мінімумів на протигагу традиційній парадигмі теорії оптимізації, яка полягає у пошуку тільки глобального мінімуму цільової функції. Ці вимоги відповідають загальній проблемі налаштування математичних моделей на поточні умови спостереження даних, що загалом відповідає визна-

ченню відповідної математичної моделі обробки даних. В даній роботі представлені результати досліджень з налаштування математичної моделі обробки даних, заснованій на критерії мінімуму протяжності [2], для випадків, коли моделлю даних є константа, а шумове оточення обумовлене шумом Коші та шумом Коші з додатними викидами й завадою у вигляді короткого гауссівського імпульсу.

**Постановка задачі та мета досліджень.** Постановка задачі полягає в дослідженні ефективності застосування математичної моделі обробки даних на основі критерію мінімуму протяжності відхилення розв'язку для складних умов спостереження даних. Метою роботи є формування рекомендацій з налаштування математичних моделей обробки даних на складне шумове оточення.

**Аналіз останніх досліджень та публікацій.** Математичні моделі обробки даних віддзеркалюють поточні умови спостереження та властивості очікуваного рішення. Вони подаються простими або складеними функціоналами з однієї, двох або більше частин, які необхідно мінімізувати або максимізувати [3]. Звичайно у якості своєї основної, а іноді й єдиної складової частини математична модель обробки даних вміщує функціонал відхилення розв'язку, який задає кількісну характеристику якості розв'язання задачі обробки даних відносно використовуваної математичної моделі даних за тим чи іншим критерієм. Традиційні математичні моделі обробки даних будуються на основі критеріїв найменших квадратів або найменших модулів, які дозволяють ефективно розв'язувати задачі обробки даних у звичайному шумовому оточенні, обумовленому шумом Гаусса або шумом Лапласа, відповідно [4]. Для більш складних шумових оточень математичні моделі обробки даних подаються на основі статистичних критеріїв, серед яких основними для практичного використання є критерії максимальної правдоподібності та узагальненої максимальної правдоподібності [1]. Застосування цих критеріїв призводить до математичних моделей обробки даних або з неопуклими функціоналами логарифмічного типу для доволі простих шумових оточень, сформованих шумом з законом узагальненого розподілу Коші [5], або з більш складними неопуклими функціоналами, побудованими на основі вартісних функцій з горизонтальними асимптотами задля запобігання впливу аномальних значень [6]. Додаткова складова частина математичної моделі обробки даних (якщо вона є доцільною до використання) вміщує функціонал, що подає властивості шуканого розв'язку або формалізує певні припущення щодо його властивостей. На даний час побудова математичної моделі обробки даних із залученням апріорних відомостей

щодо очікуваного розв'язку є звичайною практикою, оскільки це дозволяє суттєво звужити область пошуку розв'язку. Проте інколи можна вбудувати бажану властивість в основний функціонал, і тоді математична модель обробки даних спрощується, хоча проблема визначення параметру регуляризації з "зовнішньої" проблеми переходить у "внутрішню" проблему [7].

Одним з корисних для побудови математичної моделі обробки даних є критерій мінімуму протяжності [2]. Він може застосовуватися як до відхилю розв'язку, так і до шуканого розв'язку. У першому випадку він дозволяє мінімізувати кількість аномальних відхилів даних від заданої моделі, а у другому – отримувати розріджений розв'язок. Критерій мінімуму протяжності є побудованим на основі функціоналу квазіпротяжності, що для дискретного випадку має вигляд:

$$E^{(\alpha,\beta,q)}[\theta] = \sum_{n=1}^N \rho_S^{(\alpha,\beta,q)}[r_n(\theta)], \quad (1)$$

де  $r_n(\theta)$  позначає  $n$ -й елемент відхилю даних від їх параметричної моделі, заданої з точністю до вектора невідомих параметрів  $\theta$ ,  $\rho_S^{(\alpha,\beta,q)}(x) = k_S^{(\alpha,\beta,q)}[(1 + |x/\alpha|^q)^{\beta/q} - 1]$  – вартісна функція «супермножини вартісних функцій», яка залежить від трьох вільних параметрів  $\alpha$ ,  $\beta$  та  $q$ , де  $0 < \alpha < \infty$ ,  $0 < q < \infty$ ,  $-\infty < \beta \leq 1$  й  $\beta < q$ ,  $k_S^{(\alpha,\beta,q)} = 1/[(1 + |x_0/\alpha|^q)^{\beta/q} - 1]$ ,  $x_0$  – точка нормування на одиницю, тобто  $\rho_S^{(\alpha,\beta,q)}(x_0) = 1$ . Параметри  $\alpha$ ,  $\beta$ ,  $q$  мають сенс вільних параметрів, що забезпечують зміну поведінки вартісної функції та уможливають налаштування математичної моделі обробки даних на поточне шумове оточення. Функціонал квазіпротяжності (1) не є опуклим за винятком його граничних випадків, а його перша варіація в точці мінімуму може не існувати. Проте, на відміну від традиційних опуклих функціоналів, він надає можливість вирішення задач в умовах, коли дані спотворені грубими помилками, а невідомі параметри можуть приймати не одне, а декілька значень [8].

**Основна частина.** Важливою умовою отримання адекватних результатів обробки даних є її налаштування на поточне шумове оточення. Таке налаштування фактично визначає критерій обробки даних, яким доцільно користатися. Так, для випадку шумового оточення, обумовленого шумом Гаусса, найкращим для обробки є критерій найменших квадратів. У цьому випадку постановка та рішення відповідних задач обробки даних мають виконуватися в квадратичній метриці; цей та інші випадки обговорені в [4]. Далі розглянемо налаштування математичної моделі обробки даних, побудованої на основі функціоналу ква-

зіпротяжності, на просте шумове оточення, що сформоване шумом Коші, та на складне шумове оточення, що сформоване шумом Коші з додатними викидами та завадою у вигляді короткого гауссівського сигналу, за умови, що моделю даних є константа. Задаючи модель константи постійним значенням  $A_0$  та позначаючи послідовність зареєстрованих даних  $g_n$ ;  $n = 1, \dots, N$ , на основі (1) можна записати таку постановку задачі оцінювання константи:

$$\hat{A}_0 = \arg \min_{A_0} \sum_{n=1}^N \rho_S^{(\alpha, \beta, q)}(g_n - A_0). \quad (2)$$

За умови апіорно відомого значення  $A_0$ , з (2) впливає можливість налаштування її розв'язку на поточне шумове оточення шляхом вибору найкращих значень вільних параметрів. Це призводить до задачі:

$$(\hat{\alpha}, \hat{\beta}, \hat{q}) = \arg \min_{\alpha, \beta, q} \sum_{n=1}^N \rho_S^{(\alpha, \beta, q)}(g_n - A_0). \quad (3)$$

Метод розв'язання задачі (3) фактично є методом налаштування на поточне шумове оточення, який доцільно будувати на основі методів оптимізації нульового порядку для заданих послідовностей значень  $\alpha$ ,  $\beta$  та  $q$ . Нижче на рис. 1 подані приклади простого і складного шумових оточень з шумом Коші (параметр масштабу  $\nu = 0,1$ ). Рис. 2 подає результати налаштування на ці шумові оточення, а рис. 3 – динаміку налаштування за кількістю  $J$  випадкових реалізацій даних для таких послідовностей значень вільних параметрів:  $\alpha = (10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10)$ ,  $\beta = (-\infty, -16, -4, -2, -1, -1/2, -1/4, -1/8, 0, 1/8, 1/4, 1/2, 1)$ ,  $q = (1/2, 1, 3/2, 2, 5/2, 3, 4, 5, 10)$ , де @ позначає номер значення вільного параметра у його послідовності. Результати налаштування є такими: для шуму Коші  $\hat{\alpha} = \nu$ ,  $\hat{\beta} = -1/8$ ,  $\hat{q} = 2$ ; для шумового оточення на рис. 1б  $\hat{\alpha} = \nu$ ,  $\hat{\beta} = -\infty$ ,  $\hat{q} = 1,5$ . Видно, що ці результати є близькими, але не збігаються, а результат налаштування на шум Коші майже відповідає теоретичному (яким є  $\hat{\alpha} = \nu$ ,  $\hat{\beta} = 0$ ,  $\hat{q} = 2$ ) [4]. Також ці приклади вказують на природньо меншу помилку оцінювання константи під час налаштування на просте шумове оточення (рис. 3г).

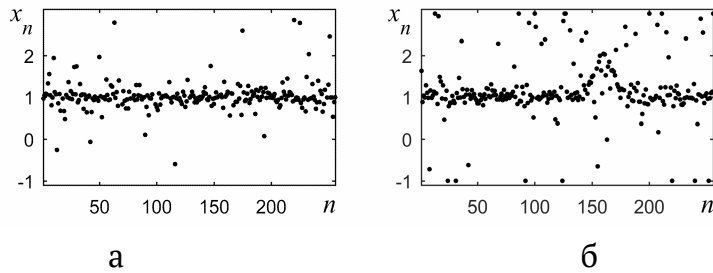


Рисунок 1 – Моделювання випадкових реалізацій постійного фону  $A_0=1$  у шумовому оточенні, обумовленому шумом Коші (а) та шумом Коші з додатними викидами й коротким гаусівським імпульсом (б)

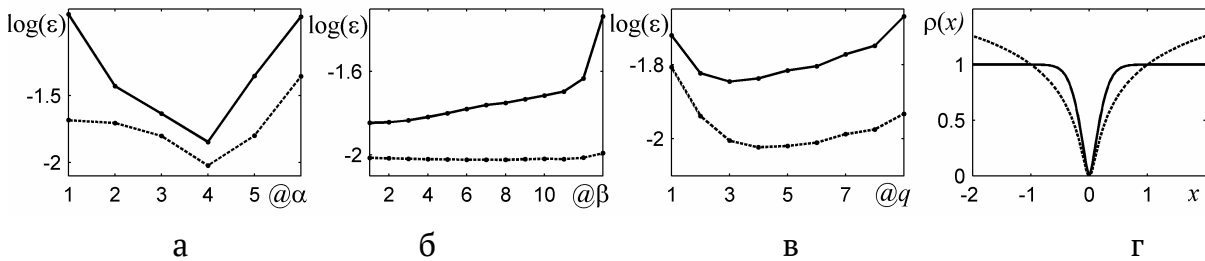


Рисунок 2 – Налаштування на шумове оточення, подане на рис.1а (нижня крива) та рис.1б (верхня крива): а, б, в – логарифм середньоквадратичної помилки для  $\alpha$ ,  $\beta$ , та  $q$ , відповідно; г – найкращі вартісні функції

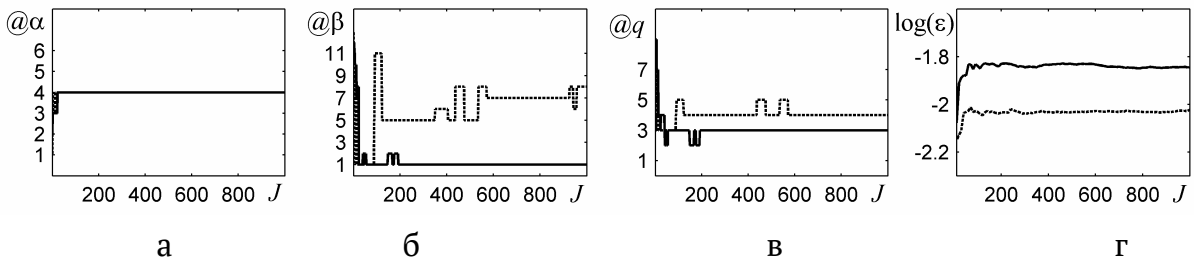


Рисунок 3 – Динаміка налаштування за числом  $J$  випадкових реалізацій шумового оточення на рис.1а (точкова крива) та рис.1б (суцільна крива): а, б, в – залежності номеру значення вільного параметру  $\alpha$ ,  $\beta$ , та  $q$ , відповідно; г – залежності логарифму середньоквадратичної помилки

Грунтуючись на результатах виконаних досліджень, можна сформулювати такі рекомендації з налаштування математичної моделі обробки даних, побудованої за критерієм мінімуму протяжності відхилю розв'язку: 1) для налаштування слід використовувати як фрагмент відомих даних з достатньою кількістю його випадкових реалізацій, так і апріорні відомості щодо шумового оточення; 2) чисельне розв'язання задачі налаштування слід виконувати методами оптимізації нульового порядку на апріорно заданій області значень вільних параметрів; 3) розв'язання задачі налаштування можна прискорити шляхом редук-

ції тривимірної задачі до двовимірної, а потім – до одновимірної через набуття вільними параметрами сталої поведінки у такій послідовності:  $\alpha$ ,  $q$ ,  $\beta$ .

**Висновки.** Ефективність застосування математичної моделі обробки даних на основі критерію мінімуму протяжності обумовлена можливістю налаштування процесу обробки даних на поточне шумове оточення. Запропонований підхід з налаштування процесу обробки даних дозволяє уникнути досліджень зі збору даних та їх статистичного опрацювання з метою визначення закону розподілу шумового оточення та є прийнятним у разі зміни шумового оточення з невідомих причин.

#### ЛІТЕРАТУРА / LITERATURE

1. Huber P. Robust statistics. 2nd ed. / P. Huber, E. M. Ronchetti . – Hoboken: Wiley. – 2009. – 370 p.
2. Вовк С. М. Критерій мінімуму протяжності / С. М. Вовк // Системні технології. Регіональний міжвузівський збірник наукових праць. – 2019. – Випуск 1 (120). – С. 19 – 25.
3. Little M. A. Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory / M. A. Little, N. S. Jones // Proceedings of the Royal Society A. – 2011. – Vol. 467. – P. 3088-3114.
4. Вовк С.М. Обробка даних за наявності шуму і грубих помилок / С. М. Вовк, В.В. Гнатушенко // Вісник Дніпропетровського університету. Серія «Фізика. Радіоелектроніка». – 2017. – Вип. 24(2). – С. 7 – 35.
5. Carrillo R. E. Generalized Cauchy distribution based robust estimation / R. E. Carrillo, T. C. Aysal, K. E. Barner // Proc. of Int. Conf. Acoustic, Speech and Signal Processing, ICASSP 2008, Las Vegas. – 2008. – P. 3389–3392.
6. Shevlyakov G. L. Robustness in data analysis: criteria and methods / G. L. Shevlyakov, N. O. Vilchevski. – Utrecht: VSP. – 2002. – 310 p.
7. Borulko V. F. Principle of minimum extent in spatial spectrum extrapolation problems of complex-valued sources / V. F. Borulko, S. M. Vovk // Telecommunications and Radio Engineering. – 2013. – Vol. 72. – N. 7. – P. 581–592. DOI: 10.1615/TelecomRadEng.v72.i7.30.
8. Вовк С.М. Оценивание параметра с несколькими значениями / С. М. Вовк, О. Н. Прокопчук // Радіоелектроніка, інформатика, управління. – 2019. – N. 4. – С. 14–24. DOI: 10.15588/1607-3274-2019-4-2.

## REFERENCES

1. Huber P. Robust statistics. 2nd ed. / P. Huber, E. M. Ronchetti . – Hoboken: Wiley. – 2009. – 370 p.
2. Vovk S. M. Kryterii minimumu protiazhnosti / S. M. Vovk // Systemni tekhnolohii. Rehionalnyi mizhvuzivskiyi zbirnyk naukovykh prats. – 2019. – Vypusk 1 (120). – С. 19 – 25.
3. Little M. A. Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory / M. A. Little, N. S. Jones // Proceedings of the Royal Society A. – 2011. – Vol. 467. – P. 3088-3114.
4. Vovk S.M. Obrobka danykh za naiavnosti shumu i hrubykh pomylok / S. M. Vovk, V.V. Hnatushenko // Visnyk Dnipropetrovskoho universytetu. Seriiia «Fizyka. Radioelektronika». – 2017. – Vyp. 24(2). – S. 7 – 35.
5. Carrillo R. E. Generalized Cauchy distribution based robust estimation / R.E. Carrillo, T. C. Aysal, K. E. Barner // Proc. of Int. Conf. Acoustic, Speech and Signal Processing, ICASSP 2008, Las Vegas. – 2008. – P. 3389–3392.
6. Shevlyakov G. L. Robustness in data analysis: criteria and methods / G. L. Shevlyakov, N. O. Vilchevski. – Utrecht: VSP. – 2002. – 310 p.
7. Borulko V. F. Principle of minimum extent in spatial spectrum extrapolation problems of complex-valued sources / V. F. Borulko, S. M. Vovk // Telecommunications and Radio Engineering. – 2013. – Vol. 72. – N. 7. – P. 581–592. DOI: 10.1615/TelecomRadEng.v72.i7.30.
8. Vovk S.M. Otsenivanie parametra s neskol'kimi znacheniyami / S. M. Vovk, O. N. Prokopchuk // Radioelektronika, Informatika, upravlinnya. – 2019. – N. 4. – S. 14–24. DOI: 10.15588/1607-3274-2019-4-2.

Received 06.11.2020.

Accepted 10.11.2020.

### **Математична модель обробки даних за критерієм мінімуму протяжності для складних умов спостереження**

*Математичну модель обробки даних для складних умов спостереження, які виникають за наявності шуму невідомої статистичної природи, потужних завад та/або раптових змін значень робочих параметрів досліджуваного об'єкта та які призводять до отримання суттєво спотворених й неповних даних, доцільно будувати на основі критерію мінімуму протяжності. Цей критерій є застосовним як до відхилю розв'язку, так і до шуканого розв'язку. У першому випадку він дозволяє мінімізувати аномальні відхилення даних від заданої моделі даних, а у другому – отримувати розріджений розв'язок. Більше того, критерій мінімуму протяжності уможливує налаштування математичної моделі обробки даних на поточне шумове оточення. Метою даної роботи є формування рекомендацій з налаштування математичних моделей обробки даних на складне шумове оточення.*

*В роботі розглянута проблема налаштування математичної моделі обробки даних на просте шумове оточення, сформоване шумом Коші, та на складне шумове оточення, сформоване шумом Коші з додатними викидами та імпульсною завадою, за умови, що моделлю даних є константа. Запропонований метод вирішення цієї проблеми реалізований як метод вирішення тривимірної задачі мінімізації функціоналу квазіпротяжності відхи-*

лу розв'язку за трьома вільними параметрами. В роботі ми демонструємо приклади значених вище простого та складного шумових оточень, які були згенеровані, результати налаштування на ці шумові оточення та поведінку налаштувань за кількістю реалізацій даних для заданих послідовностей значень вільних параметрів. Показано, що для шуму Коші результат роботи запропонованого методу налаштування збігається до теоретичного результату, а результат налаштування на шум Коші за наявності аномальних значень є близьким до попереднього за параметрами згладжування, але відрізняється від нього за параметром форми. За результатами досліджень подані рекомендації з ефективного налаштування математичної моделі обробки даних на поточне шумове оточення.

**Mathematical model of data processing based on criterion of minimum extent for complicated observation conditions**

*Under complicated observation conditions, that take place in the presence of noise of unknown statistical nature, powerful interference and/or sudden changes in the parameters values of the object under study and that lead to the significantly distorted and incomplete data, it is expedient to build a data processing mathematical model on the basis of the minimum-extent criterion. This criterion is suitable for using both to the solution residual and to the solution. In the first case, it allows to minimize the anomalous data deviations from a given data model. In the second case, it allows to obtain a sparse solution. Moreover, it allows to tune the data processing mathematical model to the current noise environment. The goal of this paper is to produce recommendations for tuning the data processing mathematical models to the current noise environment by using the minimum-extent criterion.*

*This paper considers the problem of tuning the data processing mathematical model to the simple noise environment generated by the Cauchy noise and to the complicated noise environment generated by the Cauchy noise with adding positive outliers and short pulse, provided that the data model is a constant. The proposed method for solving this problem is implemented as a method for solving the three-dimensional problem of minimizing the quasi-extent functional of the solution residual with respect to the three free parameters. In this paper, we show two examples of the mentioned above simple and complicated noise environments that have been generated, the results of tuning to these noise environments and the behaviors of tuning with respect to the number of data realizations for the given value sequences of free parameters. It is shown that for Cauchy noise the result obtained by the proposed tuning method tends to the theoretical result. However, the result obtained for the Cauchy noise with adding positive outliers and short pulse is close to the previous one by the smoothing parameters, but it differs by the parameter of the shape. According to these research, recommendations for effective tuning of data processing mathematical model to the current noise environment are presented.*

**Вовк Сергій Михайлович** - к.ф.-м.н., доцент, Дніпровський національний університет, доцент кафедри комп'ютерних наук та інформаційних технологій.

**Вовк Сергей Михайлович** - к.ф.-м.н., доцент, Днепровский национальный университет, доцент кафедры компьютерных наук и информационных технологий.

**Vovk Serhii Mikhailovic** - Ph.D, Associate Professor, Oles Honchar Dnipro National University, Associate Professor of Department of Computer Science and Information Technology.