

АДАПТАЦІЯ ЛОКАЛЬНО-ДЕТЕРМІНІСТИЧНИХ МЕТОДІВ ОПТИМІЗАЦІЇ ДО РОЗВ'ЯЗАННЯ ЗАДАЧІ ОЦІНКИ ПАРАМЕТРІВ МАТЕМАТИЧНОЇ МОДЕЛІ МЕТОДОМ ГАУСА-МАРКОВА

Гоголенко С. Ю., Святний В. А.

Донецький національний технічний університет, кафедра ЕОМ
e-mail: sergiy.gogolenko@gmail.com

Abstract

Gogolenko S. Y., Svyatnyj V. A., Adaptation of locally deterministic optimization methods for solving parameter estimation problems by means of Gauss-Markov estimator. This article contains attempt to find the most important traits of parameter estimation problem and on basis of them to propose demands' list for optimizers that solve this problem. The most suitable groups of locally deterministic optimization methods for solving parameter estimation problems are outlined and corresponding modifications for them are suggested and considered. The special attention is given to quasi-Newton optimization methods' research. Pipelining preprocessor hybrid genetic algorithms is used for search of global minimum. As a result of its research new conversion criterion was proposed.

Вступ

Стохастичні моделі станів, які отримуються як результат збудження вихідної детермінованої моделі, являють собою задачу Коші для системи стохастичних диференціальних рівнянь і описують зміни станів об'єкту, що досліджується ([2], С. 377-409). Задача параметричної ідентифікації для таких моделей має оптимізаційний характер і виникає в багатьох галузях прикладної науки та інженерії. Не зважаючи на значну кількість робіт, присвячених тематиці розробки алгоритмів оптимізації для розв'язання задачі оцінки параметрів стохастичних моделей станів ([1], [2], [7], [11], [12], тощо), в жодній з них не міститься систематизованого переліку вимог, яким повинні задовольняти подібні алгоритми. В даній статті зроблено спробу виділити важливі риси задачі оцінки параметрів з точки зору її оптимізаційного характеру та на їх основі побудувати перелік вимог до алгоритмів оптимізації, що розв'язують цю задачу. Стаття містить огляд найбільш пристосованих до виконання запропонованих вимог груп локально-детерміністичних методів (ЛДМ) оптимізації, а також ряд

пропозицій щодо їх модифікацій, спрямованих на адаптацію до розв'язання задачі оцінки параметрів. В статті наведено результати експериментальних досліджень квазіньютонівських методів, адаптованих до розв'язання задачі оцінки параметрів.

1. Особливості та характерні риси задачі оцінки параметрів моделі методом Гауса-Маркова

Найчастіше зустрічається наступний варіант постановки задачі оцінки параметрів ([11], [12]). Задано систему диференціальних рівнянь (ДР), що описує модель

$$F(t; \dot{x}(t), x(t), p) = 0, \quad x(T_0) = x_0, \quad t \in [t_0, t_0 + t_{nep}], \quad (1)$$

і набір функцій спостереження $g(x(t; x_0, p), x_0, p)$. Частина параметрів p' і початкових значень змінних стану x_0' системи ДР, що описує модель, не відомі й утворюють вектор шуканих параметрів $\theta = (x_0'; p')$. Також відома матриця замірів Y , що описує експериментальні дані. Елементи матриці Y є набором вибіркового реалізацій функцій спостереження у дискретні моменти часу $t \in T \equiv \{t_j\}_{j=1}^N \subset [t_0, t_0 + t_{nep}]$ й описуються рівнянням:

$$y_{ij} = g^{(j)}(x(t_i, \theta), \theta) + \eta_{ij}, \quad i = \overline{1, N}, \quad j = \overline{1, obs}, \quad (2)$$

де η_{ij} — деякий випадковий білий шум.

Метою є оцінка вектора шуканих параметрів, виходячи з експериментальних даних.

Для розв'язання цієї задачі зазвичай використовуються методи, що базуються на принципі максимальної правдоподібності, основною перевагою яких є інваріантність до заміни параметрів функціями над ними. В даних методах максимізується значення (або логарифм значення) густини розподілення ймовірності випадкової матриці Y , яка генерується моделлю з шуканими параметрами $\theta = (x_0'; p')$. Якщо заміри є цілком незалежними, то цільова функція, що підлягає максимізації, відповідно до методу максимальної правдоподібності приймає вигляд ([12], С.40-43)

$$f(\theta) = \sum_{i=1}^N \sum_{j=1}^{obs} \ln \pi_{\eta_{ij}}(y_{ij} - g^{(j)}(x(t_i, \theta), \theta)), \quad (3)$$

де $\pi_{\eta_{ij}}$ — густина розподілення випадкового шуму η_{ij} .

Оскільки здебільшого закон розподілення випадкової величини η_{ij} невідомий, зазвичай робиться припущення, що шум розподілений за законом Гауса з нульовим значенням математичного очікування $N(0, \sigma_{ij})$. Вибір гаусівського розподілення зумовлений тим, що будь-яке розподілення з обмеженим значенням дисперсії є асимптотично нормальним ([2], С. 355-371), а також тим, що нормальне розподілення найчастіше зустрічається у природі. Підстановка густини ймовірності за

законом нормального розподілення в формулу (3) дає метод оцінки параметрів Гауса-Маркова, згідно з яким задача оцінки параметрів приймає вигляд ([12], С. 43-44):

$$\begin{cases} \min_{\theta} \sum_{i=1}^n \sum_{j=1}^{obs} \frac{(y_{ij} - g^{(j)}(x(t_i; \theta), \theta))^2}{2\sigma_{ij}^2} \\ h(x(t), p, t) = 0 \\ g(x(t), p, t) \leq 0 \\ p^L \leq p \leq p^U \end{cases}, \quad (4)$$

де вектори p^L і p^U , а також векторні функції h і g визначають сукупність обмежень, що накладаються на параметри моделі і вектор станів.

Якщо значення дисперсії в усіх замірах однаково, то метод (4) називають незваженим методом найменших квадратів ([12], С. 44).

За відсутності систематичної помилки у моделі метод Гауса-Маркова дає асимптотично незміщену оцінку. Тому, відповідно до нерівності Рао-Крамера, нижня границя коваріаційної матриці параметрів в методі Гауса-Маркова оцінюється матрицею, зворотною до інформаційної матриці Фішера ([1], С. 149-163). Оскільки метод максимальної правдоподібності дає асимптотично R-ефективну оцінку, то при значній кількості експериментальних даних коваріаційну матрицю оцінених параметрів можна вважати рівною матриці, зворотній до інформаційної матриці Фішера. Отже, якщо число експериментальних точок велике, нелінійність моделі за параметрами, що оцінюються, не значна, помилки вимірювань η_{ij} незалежно розподілені і мають малу амплітуду, то матрицю коваріацій для шуканих параметрів можна визначити за формулою ([11]; [12], С. 250-252, С. 290):

$$P = IF^{-1}(\hat{\theta}) = E^{-1}_{y|\theta} \left\{ \left[\frac{\partial}{\partial \theta} \ln \pi_y(y|\theta) \right] \left[\frac{\partial}{\partial \theta} \ln \pi_y(y|\theta) \right]^T \right\} = -E^{-1}_{y|\theta} \left\{ \frac{\partial^2}{\partial \theta \partial \theta^T} \ln \pi_y(y|\theta) \right\} = \left(\frac{\partial^2 f(\hat{\theta})}{\partial \theta \partial \theta^T} \right)^{-1} = \left(\nabla^2 f(\hat{\theta}) \right)^{-1}, \quad (5)$$

де P — шукана матриця коваріацій, $\hat{\theta}$ — множина оцінених параметрів, $f(\theta)$ — функція максимальної правдоподібності, що оптимізується в (4).

На практиці зустрічаються різні варіанти експериментальних даних в залежності від можливості їх отримання ([2], С. 325-328). Часто зустрічається випадок, коли експериментальні дані об'єднують результати кількох випробувань. У цьому разі, якщо випробування незалежні, цільова функція є сумою цільових функцій (3) для кожного з експериментів. Матриця Фішера також розраховується шляхом підсумовування відповідних інформаційних матриць для окремих експериментів ([1], С. 164).

При використанні локально-детерміністичних методів оптимізації для оцінки параметрів виникає ряд проблем. Однією з найсуттєвіших проблем є вибір початкової точки. Локально-детерміністичні оптимізатори мають малу область простору параметрів, в яких методи сходяться до глобального мінімуму. Це впливає з того, що проблема оцінки параметрів містить значну нелінійність, а також тому, що при певних значеннях параметрів траєкторія, що є розв'язком системи модельних рівнянь, може розходитися ще до досягнення кінцевої часової точки ([11]).

Узагальнення викладеного вище матеріалу дозволяє виділити наступні характерні особливості, властиві задачі оцінки параметрів:

а) Обчислення градієнту здебільшого легко розпаралелюється. При розв'язанні задачі оцінки параметрів розрахунок цільової функції і її градієнта (або якобіана її складових) виконується одночасно з мінімальними додатковими обчислювальними витратами завдяки використанню спеціалізованих інтеграторів, що проводять одночасно інтегрування модельних рівнянь і аналіз їх чуттєвості за параметрами та початковими значеннями змінних стану.

б) У зв'язку з необхідністю інтегрування траєкторії при розрахунку значення цільової функції, її обчислення вимагає значних витрат обчислювальних ресурсів.

в) При певних значеннях параметрів можливі випадки розходження траєкторії, що є розв'язком системи модельних рівнянь. Окрім того, для багатьох систем модельних рівнянь область допустимих значень є обмеженою. Це призводить до появи прямих обмежень, які на практиці складно (або неможливо) замінити еквівалентними функціональними обмеженнями.

г) Часто при розв'язанні задачі оцінки параметрів додатково слід визначати коваріаційну матрицю, яка у випадку метода Гауса-Маркова дорівнює матриці, зворотній до гесіану цільової функції.

д) Число невідомих у даних задачах може коливатися у значних межах: від кількох змінних до кількох сотень змінних.

е) Здебільшого дані задачі погано масштабовані.

є) Цільова функція зазвичай характеризується високою нелінійністю і має численну кількість локальних мінімумів.

2. Класи локально-детерміністичних методів, пристосованих до розв'язання задачі оцінки параметрів

З перелічених вище особливостей впливає, що для ефективного розв'язання задачі оцінки параметрів методом Гауса-Маркова слід використовувати алгоритми оптимізації, орієнтовані на виконання наступних вимог:

- а) Урахування можливості одночасного обчислення градієнту і цільової функції, а також паралельного обчислення градієнту.
- б) Досягнення якомога меншого числа обчислень цільової функції.
- в) Можливість обробки прямих обмежень.
- г) Забезпечення можливості наближеного розрахунку інверсії гесіану без значних обчислювальних витрат.
- д) Можливість роботи в режимі обмеженого використання пам'яті.
- е) Інваріантність до незадовільної масштабованості цільової функції і функцій обмежень.
- є) Сполучення з методами, які забезпечують знаходження глобального мінімуму.

Серед усіх існуючих ЛДМ оптимізації (рис. 1) найбільш пристосованими до виконання наведених вимог є такі групи алгоритмів ([9], [10]):

- а) квазіньютонівські методи (ДФП, БФГШ, SR1 тощо);
- б) методи сполучених градієнтів (метод Флетчера-Рівза, метод Полака-Ріб'єра тощо);
- в) алгоритми оптимізації сум квадратів нелінійних функцій (метод Гауса-Ньютона, метод Левенберга-Маркарда тощо).

Оскільки всі ці групи алгоритмів належать до алгоритмів першого порядку, в них автоматично враховується можливість одночасного обчислення градієнту і цільової функції. Квазіньютонівські методи і методи сполучених градієнтів мають зверхлінійну швидкість сходження, а методи оптимізації сум квадратів нелінійних функцій, такі як метод Гауса-Ньютона і метод Левенберга-Маркарда, показують квадратичну швидкість сходження для задач з нульовою нев'язкою, тому в них під час розв'язання оптимізаційних задач виконується відносно незначна кількість обчислень цільової функції у порівнянні з іншими методами нульового та першого порядків.

Класичні реалізації зазначених груп методів не орієнтовані на розв'язання задач з прямими обмеженнями. Усунути цей недолік можна при використанні моделі лінійного пошуку. Для цього достатньо в алгоритмах пошуку крокового множника модифікувати процедуру обору пробних значень таким чином, щоб при потраплянні поточного пробного значення в зону прямих обмежень генерувалося нове пробне значення, ймовірність розташування якого у множині допустимих значень була більшою. Основним недоліком такого підходу є неврахування можливої відсутності підмножин, що задовольняють на правила вибору крокового множника (правила Вулфа, правила Голдштейна тощо) ([10], С. 36-41).

Квазіньютонівські методи організовані таким чином, що вони, базуючись на грубій оцінці матриці Гесе в початковій точці, поступово будують точнішу апроксимацію гесіану, використовуючи інформацію про градієнт функції, отриману на кількох або всіх попередніх ітераціях. У

методах Гауса-Ньютона і Левенберга-Маркарда також обчислюються складові, необхідні для наближеного розрахунку значення гесіану цільової функції. Тому визначення апроксимації інверсії гесіану в квазіньютонівських методах і методах оптимізації сум квадратів нелінійних функцій зазвичай не вимагає значних додаткових обчислювальних витрат.

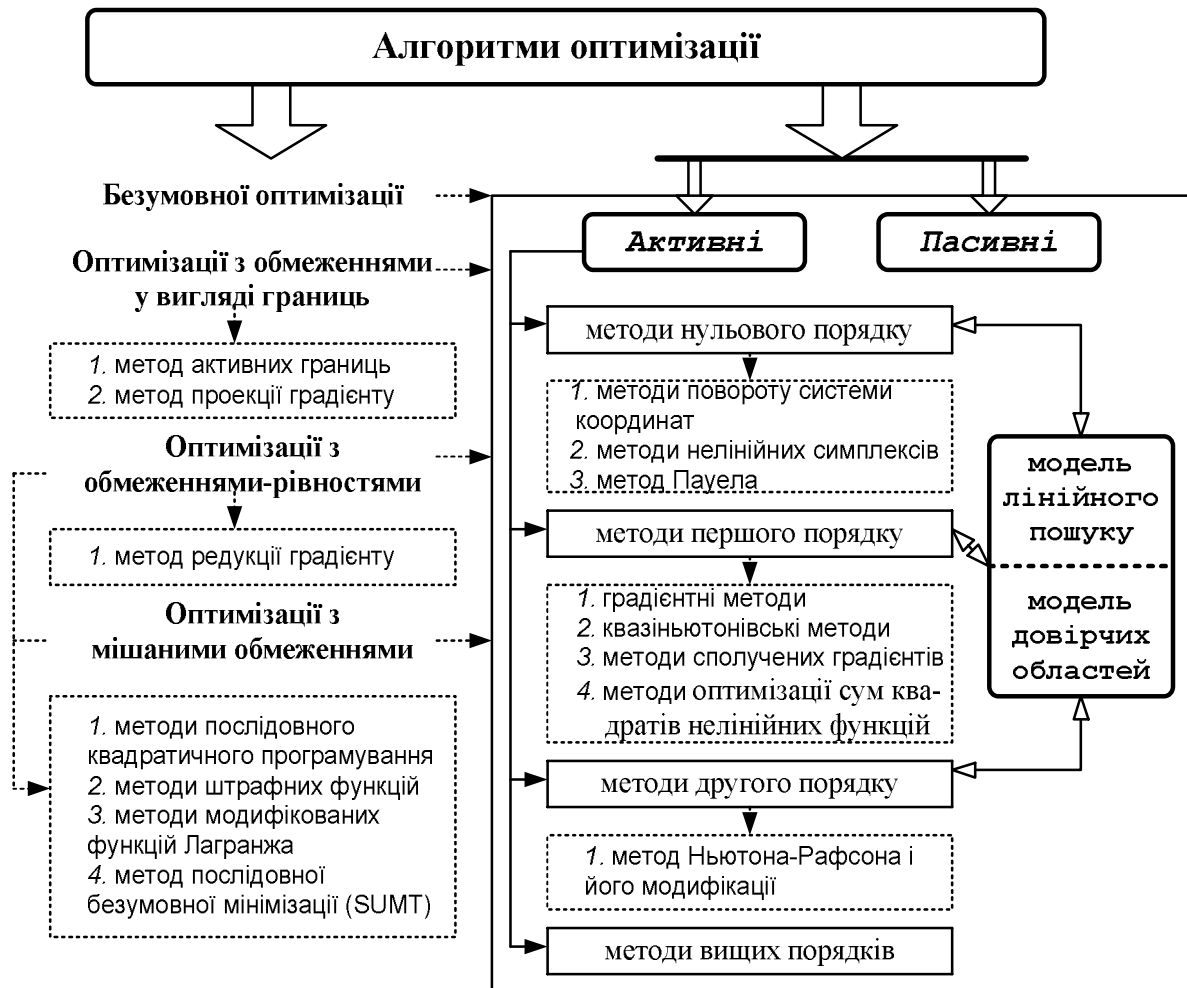


Рисунок 1. Класифікація локально-детерміністичних методів нелінійного програмування

Основним недоліком методів оптимізації сум квадратів нелінійних функцій у порівнянні з квазіньютонівськими методами та методами сполучених градієнтів є значний ріст необхідних об'ємів машинної пам'яті зі зростанням розмірності задачі. У квазіньютонівських методах дана проблема розв'язується за допомогою використання спеціалізованих квазіньютонівських матриць з обмеженим використанням пам'яті ([10], С. 222-247).

Додатковою перевагою квазіньютонівських методів перед іншими групами алгоритмів є досягнення ними інваріантності до поганої

масштабованості цільової функції при виконанні певної кількості кроків ([9], [10]).

Одним з основних підходів до розв'язання проблеми вибору початкового значення для локально-детерміністичного методу є підхід з оцінкою початкової точки (initial-value approach) ([7], С. 9). Він полягає в тому, що після належного вибору початкового наближення модельні рівняння інтегруються на всьому часовому проміжку оцінки параметрів, а потім обчислюється цільова функція задачі (4). Даний підхід вимагає використання на початковому етапі методу, що забезпечує грубе наближення до глобального мінімуму. Найпростішим виходом із цієї ситуації є використання деякого глобального стохастичного оптимізатора. У даній роботі в якості такого обрано генетичний алгоритм оптимізації ([3], [5]). Єдиним недоліком є те, що підхід з оцінкою початкової точки стохастичним оптимізатором не розв'язує проблеми розходження розв'язку системи модельних рівнянь.

3. Реалізація підходу з оцінкою початкової точки шляхом застосування гібридних генетичних алгоритмів (ГА)

Алгоритми оптимізації, в яких ГА і ЛДМ працюють послідовно, тобто перший генерує дані (наприклад, початкові наближення) для другого, належать до групи зчеплених препроцесорних гібридних ГА (рис.2). Здебільшого ГА в зчеплених гібридних алгоритмах використовується для виділення початкового простору пошуку таким чином, щоб ЛДМ схилювався швидко і з високою точністю ([15]).

Перехід між ГА і ЛДМ в зчеплених гібридних алгоритмах здебільшого здійснюється за наперед заданими критеріями. Існуючі критерії переходу умовно поділяються на три групи:

- а) критерії, в яких аналізуються параметри генетичного алгоритму;
- б) критерії, в яких перевіряється відносна однорідність популяції;
- в) критерії, що використовують додаткові знання про задачу.

Одним з параметрів ГА, аналіз якого дозволяє зробити евристичні висновки про доцільність переходу до ЛДМ, виступає процент формування схеми. Граничним випадком при цьому є перехід від ГА до ЛДМ, коли схема повністю сформована, тобто ГА завершив своє виконання. Використання даного параметру доцільно комбінувати з кодуванням за Греєм в ГА.

В критеріях другої групи певним чином проводиться аналіз топології пробних точок, які формують поточну популяцію. Можливо, найпростішим способом перевірки однорідності популяції, тобто скопичування точок навколо одного мінімуму, є аналіз значення коефіцієнту варіації ([6]). Складніші підходи полягають у використанні регресійного аналізу ([6]). В них уся область пошуку поділяється на

підобласті, що утворюють загальну регресійну модель. Кожній з підобластей ставиться у відповідність поліноміальна (чи криволінійна) модель. Найчастіше регресійну модель складає множина лінійних чи квадратичних моделей. На кожній з ітерацій ГА виконується підгонка параметрів моделі до цільової функції за значеннями точок, які формують поточну популяцію, а потім за отриманими даними розраховується коефіцієнт множинної кореляції (коефіцієнт детермінації). Якщо отримане значення коефіцієнту множинної кореляції достатньо високе, то виконується перехід до ЛДМ. В якості початкового наближення для ЛДМ здебільшого обирається точка глобального мінімуму регресійної моделі.

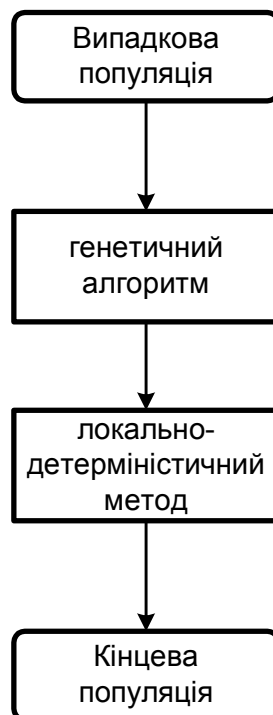


Рисунок 2. Архітектура зчепленого препроцесорного гібридного ГА

До третьої групи критеріїв можна віднести перевірку значення градієнту цільової функції. Якщо градієнт більший за певне наперед задане число, яке вибирається відповідно до масштабованості цільової функції або з інших міркувань, то за даним критерієм відбувається перехід до ЛДМ.

4. Експериментальні дослідження

З метою визначення оптимального для розв'язання задачі оцінки параметрів настроювання алгоритмів оптимізації було проведено ряд експериментальних досліджень. В якості базової для досліджень було обрано групу квазіньютонівських методів. Розглядалися технічні рішення, запропоновані в багатьох сучасних пакетах оптимізації, таких як LBFGS-B,

GAUSS, TNPACK, BTN, MINPACK-2, SNOPT та PROC NLP ([9], [13]). Зокрема серед процедур вибору довжини кроку тестуванню підлягали проста процедура бектрекінгу, процедура бектрекінгу Полака та двохфазові процедури, побудовані за схемою S2 Ал-Баалі-Флетчера, схемою Ноцедаля і схемою Мора-Соренсена. У випадку двохфазових процедур при виборі пробних значень використовувалися алгоритми близькі до алгоритму Мора і Туенте ([8]). Для оновлення, квазіньютонівської матриці в тестах використовувалися формулиДФП, БФГШ і SR1. Також розглядався варіант, коли квазіньютонівська матриця на кожному кроці лишалася незмінною діагональною матрицею (градієнтний метод). Для вибору початкової точки використовувався зчеплений препроцесорний гібридний ГА з критерієм переходу за відсотком формування схеми.

Множина задач, що використовувалася для тестування реалізованих в квазіньютонівському оптимізаторі процедур вибору крокового множника, включала функції Янаї-Озава-Канеко, Плассмана і Мора ([8], [14]). Тестування гібридного ГА проводилося на функціях Аклі, Діксона, Розенброка, Швевеля, Шеллоу і Шенно ([3], [4]). В результаті тестування на перелічених штучних цільових функціях було визначено, що найбільш ефективною є реалізація гібридного ГА, в якому в квазіньютонівському алгоритмі для побудови апроксимації гесіану використовувалася формула оновлення БФГШ ([10], С. 194-201, С. 211-218), а для пошуку довжини кроку — схема Мора-Соренсена ([8]).

Окрім штучних цільових функцій, дослідження алгоритму проводилися також на задачі оцінки параметрів моделі Лотки-Вольтерра (Lotka-Volterra). Дана модель, відома також як модель здобичі-хижака, описує динаміку біологічних систем, в яких розвиваються два різновиди біологічних істот — хижаки та їх здобич. Вона має достатньо просту структуру і описується системою з двох нелінійних диференціальних рівнянь (ДР):

$$\begin{cases} \frac{dx}{dt} = x(p_1 - p_2 y) \\ \frac{dy}{dt} = -y(p_3 - p_4 x) \end{cases}, \quad (6)$$

де x — кількість жертв, y — кількість хижаків, p_1 , p_2 , p_3 і p_4 — параметри моделі.

Для проведення досліджень генерувалися тестові файли даних, в яких наводилися значення величин x і y у ста точках часового проміжку від нуля до десяти одиниць часу. Шум даних складав від нуля до трьох відсотків. Оцінці підлягали усі чотири параметри моделі. В усіх тестах при виборі початкової точки в області опуклості поблизу розв'язку квазіньютонівський алгоритм задовільно знаходив значення шуканих параметрів. Тести показали, що досягнення глобального мінімуму

гібридним алгоритмом з достатньо високою ймовірністю можна досягти лише при зупинці ГА з процентом формування схеми не менше 99,8%. В той же час за результатами тестування можна стверджувати, що зі зростанням номеру ітерації мінімальний необхідний для переходу до квазіньютонівського алгоритму процент формування схеми доцільно зменшувати. Це впливає, зокрема, з того, що в останніх ітераціях ГА значення цільової функції майже не змінювалося (рис.3,а).

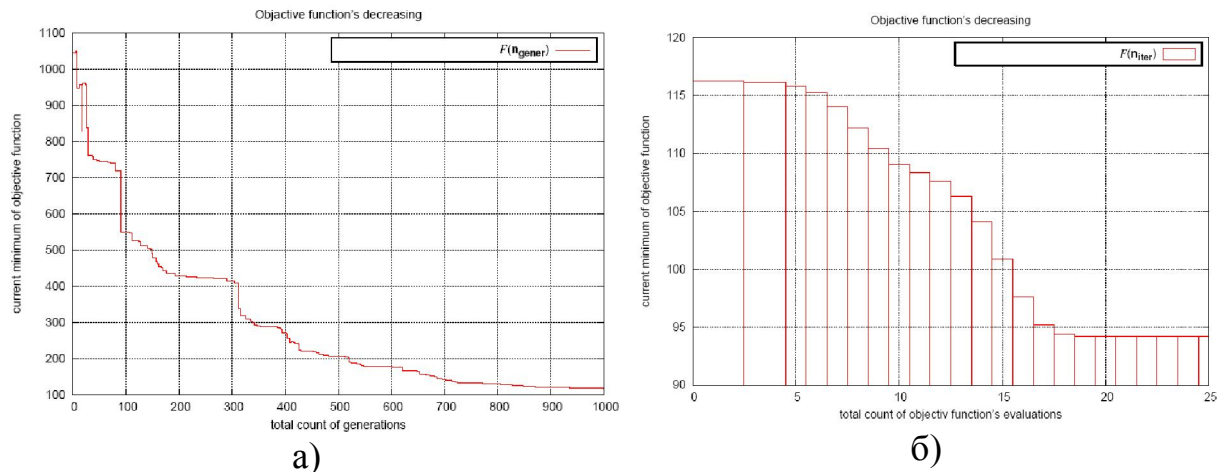


Рисунок 3. Зменшення цільової функції при параметричній ідентифікації моделі Лотки-Вольterra в залежності від числа генерацій ГА (а) і кількості обчислень цільової функції квазіньютонівським алгоритмом (б)

Висновки

В даній статті проведено аналіз задачі оцінки параметрів математичної моделі, на основі якого запропоновано групу критеріїв до побудови алгоритмів оптимізації, пристосованих до розв'язання цієї задачі. Базуючись на запропонованих критеріях, було виділено та досліджено класи ЛДМ, найбільш пристосованих до розв'язання задачі оцінки параметрів. В статті також розглянуто реалізацію підходу до розв'язання задачі параметричної ідентифікації з оцінкою початкової точки шляхом застосування зчеплених препроцесорних гібридних ГА. В результаті дослідження роботи зчеплених гібридних ГА було запропоновано модифікацію критерію переходу від ГА до ЛДМ за значенням проценту формування схеми з динамічним контролем даного параметру. Ефективність та математичний аналіз запропонованого критерію підлягає подальшому дослідженню.

Література

1. Боровков А.А. *Математическая статистика: оценка параметров, проверка гипотез.* — М.: «Наука», 1984 — 472 с.
2. Волков И.К., Зуев С.М., Цветкова Г.М. *Случайные процессы: учеб. для вузов.* — М.: Изд-во МГТУ им. Н.С. Баумана, 1999. — 448 с.
3. Ваек Th. *Evolutionary Algorithms in Theory and Practice.* — New York: Oxford University Press, Inc., 1996.
4. Bassiri K., Hutchinson D. *New Parallel Conjugate Directions Method for Unconstrained Optimisation* // School of Computer Studies Report Series, Technical Report 94.2. — Leeds: University of Leeds, 1996. — 46 с.
5. Cantu-Paz E. *Efficient and accurate parallel genetic algorithms.* — Boston, Dordrecht, London: Kluwer Academic Publishers, 2001. — 162 с.
6. Eddy J., Hacker K., Lewis K. *Tuning a hybrid optimization algorithm by determining the modality of the design space* // ASME 2001 Design Engineering Technical Conferences and Computers and Information in Engineering Conference. — Pittsburgh, 2001. — 10 с.
7. Horbelt W. *Maximum likelihood estimation in dynamical systems.* — Freiburg: Albert-Ludwigs-Universitaet Freiburg Fakultat fuer Physik, 2001.
8. More J., Thuente D. *Line Search Algorithms with Guaranteed Sufficient Decrease* // ACM Transactions on Mathematical Software, 1994. — Т. 20, No. 3. — С. 286–307.
9. More J., Wright S. *Optimization software guide.* — Philadelphia: SIAM Soc. for Industrial and Applied Mathematics, 1994. — 154 с.
10. Nocedal J., Wright S. *Numerical optimization.* — New York: Springer, Springer series in operations research, 1999. — 623 с.
11. Peifer M., Timmer, J. *Parameter estimation in ordinary differential equations using the method of multiple shooting — a review.* — Freiburg: Freiburg Centre for Data Analysis and Modelling, 2005. — 21 с.
12. Walter E., Pronzato L. *Identification of parametric models from experimental data.* — Paris, Milan, Barcelona: Springer, 1997. — 403 с.
13. Wright S. *Optimization Software Packages* // Handbook of Applied Optimization, M. Resende and P. Pardalos, eds. — Oxford: Oxford University Press, 2002. — С. 1008-1015.
14. Yanai H., Ozawa M., Kaneko S. *Interpolation methods in one dimensional optimization* // Computing, — Wien: Springer, 1981 — Т.27, No.2 — С. 151–163.
15. Yen J., Liao J., Lee B., Randolph D. *A Hybrid Approach to Modeling Metabolic Systems Using Genetic Algorithm and Simplex Method* // 11th Conference on Artificial Intelligence for Applications. — 1995. — 42 с.