

## Секція 4. Паралельні та розподілені обчислення

УДК 004.9

### РЕАЛІЗАЦІЯ ГУСЕНИЦЯ-SSA ДЛЯ ПАРАЛЕЛЬНИХ СИСТЕМ З РОЗПОДІЛЕНОЮ ПАМ'ЯТТЮ

**Белова І.В., Назаревич О.Б.**

НТУУ «Київський політехнічний інститут»

Тернопільський національний технічний університет імені Івана Пулюя

#### **I. Постановка проблеми**

В основі роботи лежить аналіз стану газових мереж при реагуванні на аварії систем газопостачання та виявленні несанкціонованого відбору газу. Процес газоспоживання розглядаємо як випадковий процес з невідомими параметрами. Для виділення складових газопостачання, які будуть характеризувати обсяги несанкціонованого відбору газу, запропоновано використати метод «Гусениця-SSA». Оскільки плануємо використовувати даний метод у реальному часі оцінки процесу газоспоживання і виявляти параметри випадкового процесу, то необхідно приділити увагу обчислювальній складності задачі. Аналіз складових газопостачання в реальному часі дозволить знизити непродуктивні витрати газу і оперативно реагувати на проблеми.

#### **II. Мета роботи**

Метою дослідження є розробка програмного забезпечення для швидкісного виділення складових газопостачання за допомогою методу Гусениця-SSA. Так як існуючі реалізації для систем зі спільною пам'яттю не мають такого рівня масштабування, який би дозволив досягти необхідної швидкості розрахунків, то було виконано відповідну реалізацію для систем з розподіленою пам'яттю. За допомогою попереднього аналізу предметної області було підвищено і швидкість виконання й локалізації методу під поставлену задачу.

#### **III. Особливості програмної реалізації методу «Гусениця-SSA» для паралельних систем з розподіленою пам'яттю**

Для високопродуктивної реалізації даного дослідження було використано кластер суперкомп'ютерного центру НТУУ «КПІ», 13 вузлів, 2x Intel Xeon 5160, 3 ГГц, 4Гб RAM. Мовою застосування стала C, OpenMPI 1.3.3, MKL 10.1 (зокрема ScaLAPACK), MPI FFTW.

У процесі аналізу предметної галузі ми виявили, що для поставлених цілей достатньою буде одинарна точність обчислень з плаваючою комою.

Найбільш трудомістким етапом методу «Гусениця-SSA» з обчислювальної точки зору є момент обчислення SVD розкладу матриці великої розмірності. Для цього було включено метод ScaLAPACK `psgesvd_`, який розподіляє матрицю по процесорам за допомогою двовимірного циклічного розподілу й обчислює значення сингулярних чисел і векторів для заданої матриці.

На етапі групування було використано різноманітні підходи для прискорення обчислень. При аналізі предметної області визначено ті частоти, на яких у даній задачі необхідно шукати компоненти розкладу. Також, опираючись на існуючі дані, було обчислено ті значення сингулярного розкладу, які несуть основний вплив на значення компонент. При групуванні:

- Для значень лівого сингулярного вектора за допомогою MPI FFTW функції `fftwf_plan_many_dft_r2c` рахуємо частотні складові, з яких відшукуємо максимум. Він характеризує максимальну амплітуду, яка і буде задавати поведінку складової.
- Для підрахунку максимальних частотних складових додатково виконуємо паралелізацію лівого сингулярного вектора на процесори.
- Шукаємо індекси, які відповідають обраним компонентам розкладу.
- Обчислюємо траєкторні матриці, які відповідають визначеним індексам. Паралелізацію обчислення виконуємо за допомогою PBLAS функції `psger_`, яка розподіляє матрицю по процесорам за допомогою двовимірного циклічного розподілу. При цьому виконуємо матрично-векторну операцію множення власних векторів і додавання складових матриць. Збір даних із процесорів проводимо лише при отриманні результуючої матриці компоненти.

При діагональному усередненні паралелізація була виконана для отримання результуючого вектора траєкторії компоненти завдяки паралельному усередненню незалежних елементів цього вектору.

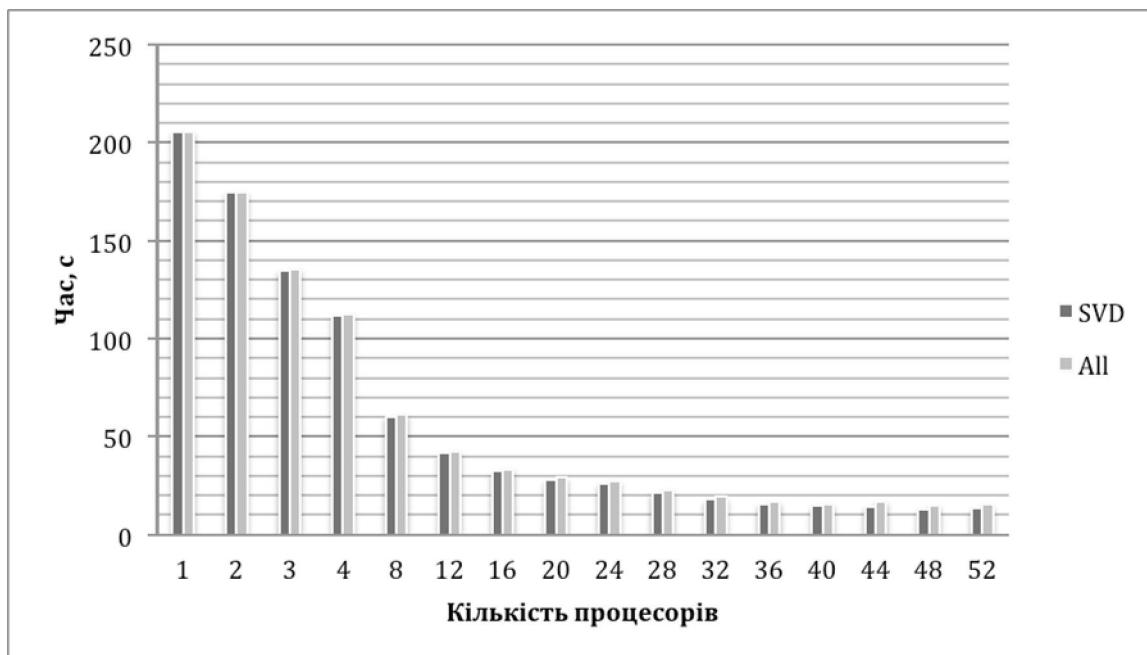


Рисунок 1 – Експериментальні оцінки продуктивності реалізації паралельного методу «Гусениця-SSA» для систем із розподіленою пам'яттю з метою виділення тренду часового ряду газопостачання.

Сумарна швидкість реалізації «Гусениці-SSA» для систем із паралельною пам'яттю та прив'язкою до предметної області у декілька разів перевищила швидкість роботи існуючих реалізацій для систем зі спільною пам'яттю на одному процесорі, та на порядок - при виконанні цього методу на 32 процесорах. Прискорення роботи методу зображено на рисунку 1.

#### Список використаних джерел

1. Golyandina N. E., Nekrutkin V. V., Zhigljavsky A. A. Analysis of Time Series Structure: SSA and related technique — Chapman & Hall / CRS, Boca Raton, 2001.
2. Blackford L. S., Choi J., Cleary A., D'Azevedo E., Demmel J., Dhillon I., Dongarra J., Hammarling S., Henry G., Petitet A., Stanley K., Walker D., Whaley, R. C. ScaLAPACK Users' Guide - Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997
3. Matteo Frigo and Steven G. Johnson, "The Design and Implementation of FFTW3," Proceedings of the IEEE 93 (2), 216–231, 2005

УДК 681.3.01

## МОДЕЛЬ ЕКОНОМІЧНО-ЕФЕКТИВНОГО ЗАСТОСУВАННЯ УКРАЇНСЬКОГО НАЦІОНАЛЬНОГО ГРІД ДЛЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ З МАЛИМ ЧАСОМ ОБЧИСЛЮВАЛЬНОЇ ТРАНЗАКЦІЇ

Ільїн М.І., Ільїн К.І.

НТУУ «Київський політехнічний інститут»

На практиці часто зустрічаються задачі, обчислювальна складність яких обумовлена необхідністю багатократного виконання відносно швидкої функції (задачі з природнім паралелізмом, з розділенням за даними). Серед прикладів можна навести медичні популяційні дослідження електрокардіограм [1] та прогнозування наслідків впливу бюджетно-податкової політики держави на економіку регіонів [2]. Наведені приклади та багато інших задач мають додаткову властивість – рідкість необхідності проведення обчислень, а саме в [2] необхідність у моделюванні виникає один раз на рік (під час прийняття рішень в ситуаційному центрі в рамках обговорення держбюджету