Журнал обчислювальної та прикладної математики | 2022, № 1 | Journal of Numerical & Applied Mathematics

УДК 519.85

MSC 37C75, 65K05

CONVERGENCE OF GRADIENT-LIKE DYNAMICAL SYSTEM

A. Yu. Shavlyuk, V. V. Semenov

Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, E-mail: shavliuk777@gmail.com, volodya.semenov@gmail.com

ЗБІЖНІСТЬ ГРАДІЄНТНОЇ ДИНАМІЧНОЇ СИСТЕМИ

А. Ю. Шавлюк, В. В. Семенов

Факультет комп'ютерних наук та кібернетики, КНУ імені Тараса Шевченка, Київ, Україна, E-mail: shavliuk777@gmail.com, volodya.semenov@gmail.com

Abstract. The asymptotic behavior of the gradient system, which is a continuous analogue of the variant of the gradient method from [16] for the minimization of strongly convex functions, is studied. Using the Lyapunov analysis, estimates of the rate of convergence of the gradient system were established.
Keywords: convexity, optimization, gradient system, convergence, Lyapunov function.

Анотація. Досліджено асимптотичну поведінку градієнтної системи, яка є неперервним аналогом варіанту градієнтного методу з [16] для мінімізації сильно опуклих функцій. За допомогою другого методу Ляпунова встановлено оцінки швидкості збіжності градієнтної системи до точки рівноваги.
Ключові слова: опуклість, оптимізація, градієнтна система, збіжність, функція Ляпунова.

Вступ

З гладкою задачею мінімізації

$$f \to \min$$

можна зв'язати градієнтну систему

$$\begin{cases} \dot{x} = -\nabla f(x), \\ x(0) = x_0. \end{cases} \tag{1}$$

Класичний метод градієнтного спуску є результатом дискретизації задачі Коші (1) за допомогою явного методу Ейлера ($h > 0$):

$$\frac{x_{k+1} - x_k}{h} = -\nabla f(x_k).$$

За різних умов у багатьох роботах [1–5] за допомогою другого методу Ляпунова доведено, що траєкторії системи (1) при великих $t > 0$ притягуються до множини $S = \{x : \ \nabla f(x) = 0\}$ точок рівноваги системи (1).

Існують й інші диференціальні рівняння, траєкторії яких мінімізуючі. Наприклад, так званий метод важкої кульки Поляка [4, 6–8]:

$$\begin{cases} \ddot{x} + a\dot{x} = -\nabla f(x), \\ x(0) = x_0, \ \ \dot{x}(0) = 0, \ \ a > 0; \end{cases} \qquad \begin{cases} \dot{x} = y, \\ \dot{y} = -ay - \nabla f(x), \\ x(0) = x_0, \ \ y(0) = 0, \ \ a > 0. \end{cases}$$

У роботі [9] розглянуто задачу Коші:

$$\begin{cases} \ddot{x} + \frac{3}{t}\dot{x} = -\nabla f(x), \\ x(t_0) = x_0, \ \ \dot{x}(t_0) = 0, \ \ t_0 > 0; \end{cases} \qquad \begin{cases} \dot{x} = y, \\ \dot{y} = -\frac{3}{t}y - \nabla f(x), \\ x(t_0) = x_0, \ \ y(t_0) = 0, \ \ t_0 > 0. \end{cases} \tag{2}$$

Якщо здійснити дискретизацію задачі Коші (2) таким чином

$$\frac{1}{h^2}\left(x_{k+1} - 2x_k + x_{k-1}\right) + \frac{3}{kh^2}\left(x_k - x_{k-1}\right) = -\nabla f(y_k),$$

де $y_k = x_k + \left(1 - \frac{3}{k}\right)\left(x_k - x_{k-1}\right)$, то поклавши $\lambda = h^2$, приходимо до відомого швидкого градієнтного методу Нестерова [10]:

$$\begin{cases} y_k = x_k + \frac{k-3}{k}\left(x_k - x_{k-1}\right), \\ x_{k+1} = y_k - \lambda \nabla f(y_k). \end{cases}$$

В роботі [11] для $\alpha \geq 3$ розглянуто задачі Коші:

$$\begin{cases} \ddot{x} + \frac{\alpha}{t}\dot{x} = -\nabla f(x), \\ x(t_0) = x_0, \ \ \dot{x}(t_0) = 0, \ \ t_0 > 0; \end{cases} \qquad \begin{cases} \dot{x} = y, \\ \dot{y} = -\frac{\alpha}{t}y - \nabla f(x), \\ x(t_0) = x_0, \ \ y(t_0) = 0, \ \ t_0 > 0. \end{cases} \tag{3}$$

Задача Коші (3) майже не відрізняється від (2), але має кращу асимптотичну поведінку при $\alpha > 3$.

Зауважимо, що неперервні методи мінімізації (градієнтні системи) привабливі тим, що для їх інтегрування можна використати не тільки явний метод Ейлера, але й інші методи, які збігаються швидше та краще пристосовані для інтегрування жорстких систем диференціальних рівнянь. До жорстких градієнтних систем приходимо при мінімізації яружних функцій.

На цьому шляху можна отримати алгоритми мінімізації, які важко знайти, залишаючись в рамках звичних уявлень про дискретні ітераційні процеси. Це стимулює розвиток неперервних методів розв'язання екстремальних та близьких задач [12–15].

В даній статті досліджується асимптотична поведінка градієнтної системи, яка є неперервним аналогом варіанту градієнтного методу з [16] для мінімізації сильно опуклих функцій. За допомогою другого методу Ляпунова встановлено оцінки швидкості збіжності градієнтної системи до точки рівноваги.

1. Градієнтна система

Нагадаємо, що функцію $f : \mathbb{R}^n \to \mathbb{R}$ називають L-гладкою, якщо вона диференційовна та її градієнт задовольняє умову Ліпшиця з константою $L > 0$ [10]. Для L-гладкої функції має місце нерівність [10]

$$f(y) \le f(x) + (\nabla f(x), y - x) + \frac{L}{2}\|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n.$$

Функцію $f : \mathbb{R}^n \to \mathbb{R}$ називають сильно опуклою з константою $\mu > 0$, якщо

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - \frac{1}{2}\lambda(1-\lambda)\mu\|x - y\|^2$$

для всіх x, $y \in \mathbb{R}^n$, $\lambda \in [0,1]$. Сильно опукла функція $f : \mathbb{R}^n \to \mathbb{R}$ досягає мінімуму в єдиній точці x_*, причому

$$f(x) - f(x_*) \ge \frac{\mu}{2}\|x - x_*\|^2 \quad \forall x \in \mathbb{R}^n.$$

Для диференційовної сильно опуклої функції f має місце нерівність [10]

$$f(y) \ge f(x) + (\nabla f(x), y - x) + \frac{\mu}{2}\|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n.$$

Для L-гладкої сильно опуклої функції $f \in C^2(\mathbb{R}^n)$ друга похідна додатньо визначена [10], точніше для всіх $x \in \mathbb{R}^n$

$$L\|\xi\|^2 \ge (\nabla^2 f(x)\xi, \xi) \ge \mu\|\xi\|^2 \quad \forall \xi \in \mathbb{R}^n.$$

Припустимо, що C^2-функція $f : \mathbb{R}^n \to \mathbb{R}$ сильно опукла (з константою $\mu > 0$) та L-гладка. Нехай

$$f_* = f(x_*) = \min_{x \in \mathbb{R}^n} f(x).$$

Розглянемо градієнтну систему вигляду:

$$\begin{cases} \dot{x}(t) = -d(x(t))\nabla f(x(t)), \\ d(x(t)) = \frac{\|\nabla f(x(t))\|^2}{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))}, \\ x(0) = x_0. \end{cases} \tag{4}$$

Використовуючи для наближеного інтегрування задачі Коші (4) явну схему Ейлера, отримуємо такий алгоритм:

$$x_{k+1} = x_k - \tau_k \frac{\|\nabla f(x_k)\|^2}{(\nabla^2 f(x_k)\nabla f(x_k), \nabla f(x_k))}\nabla f(x_k), \ \ \tau_k > 0. \tag{5}$$

Зауваження 1. Алгоритм (5) (з $\tau_k \equiv 1$) запропоновано в [16] для розв'язання деяких задач синтезу оберненого звязку в лінійних динамічних системах. Логіка схеми така. У формулі $x_{k+1} = x_k - \lambda\nabla f(x_k)$ слід обрати множник $\lambda > 0$. Розглянемо функцію $\varphi(t) = f(x_k - t\nabla f(x_k))$, $t \in [0, +\infty)$, та робимо один крок методу Ньютона

$$t_1 = t_0 - \frac{\varphi'(t_0)}{\varphi''(t_0)}, \quad t_0 = 0.$$

Отримуємо

$$\lambda = t_1 = \frac{\|\nabla f(x_k)\|^2}{(\nabla^2 f(x_k)\nabla f(x_k), \nabla f(x_k))}.$$

Зауваження 2. У монографії [17] для пошуку нулів диференційовних операторів F, що діють в гільбертовому просторі H, розглядався близький до (5) ітераційний процес

$$x_{k+1} = x_k - \frac{\|Fx_k\|^2}{(F'(x_k)Fx_k, Fx_k)} Fx_k. \tag{6}$$

Ясно, що

$$\frac{1}{L} \le d(x(t)) = \frac{\|\nabla f(x(t))\|^2}{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))} \le \frac{1}{\mu}$$

та задача Коші (4) має єдиний розв'язок $x : [0, +\infty) \to \mathbb{R}^n$. Система (4) має єдину точку рівноваги — точку мінімуму функції f.

Перейдемо до обгрунтування схеми (4), тобто доведемо асимптотичну стійкість (4).

2. Ляпуновський аналіз

Проведемо аналіз асимптотичних властивостей градієнтної системи (4) за допомогою другого методу Ляпунова.

Лема 1. *Має місце тотожність*

$$\|\nabla f(x(t))\| = e^{-t}\|\nabla f(x_0)\| \quad \forall t > 0. \tag{7}$$

Доведення. Розглянемо

$$\tfrac{d}{dt}\left(\tfrac{1}{2}\|\nabla f(x(t))\|^2\right).$$

Маємо,

$$\begin{aligned}\tfrac{d}{dt}\left(\tfrac{1}{2}\|\nabla f(x(t))\|^2\right) &= (\nabla^2 f(x(t))\dot{x}(t), \nabla f(x(t))) = \\ &= -d(x(t))(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t))) = \\ &= -\frac{\|\nabla f(x(t))\|^2}{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))}(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t))).\end{aligned}$$

Звідки,

$$\tfrac{d}{dt}\left(\|\nabla f(x(t))\|^2\right) + 2\|\nabla f(x(t))\|^2 = 0,$$

тобто,

$$\tfrac{d}{dt}\left(e^{2t}\|\nabla f(x(t))\|^2\right) = 0.$$

Отже,

$$e^{2t}\|\nabla f(x(t))\|^2 = \|\nabla f(x_0)\|^2 \quad \forall t > 0.$$

Тотожність (7) доведена. □

Лема 2. *Мають місце нерівності*

$$\|x(t) - x_*\| \le e^{-\frac{\mu}{2L}t}\|x_0 - x_*\| \quad \forall t > 0, \tag{8}$$

$$f(x(t)) - f_* \le \sqrt{\tfrac{\mu}{L}} e^{-\frac{\mu}{2L}t}\left(f(x_0) - f_*\right) \quad \forall t > 0. \tag{9}$$

Доведення. Розглянемо функцію

$$V(t) = \tfrac{1}{2}\|x(t) - x_*\|^2.$$

Знайдемо $\frac{d}{dt}V(t)$. Маємо,

$$\begin{aligned}
\tfrac{d}{dt}V(t) &= (x(t) - x_*, \dot{x}(t)) = \\
&= \frac{\|\nabla f(x(t))\|^2}{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))}(x_* - x(t), \nabla f(x(t)) \le \\
&\le \frac{\|\nabla f(x(t))\|^2}{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))}\left(f_* - f(x(t)) - \tfrac{\mu}{2}\|x(t) - x_*\|^2\right) \le \\
&\le -\tfrac{\mu}{2L}\|x(t) - x_*\|^2 = -\tfrac{\mu}{L}V(t).
\end{aligned}$$

Звідки,

$$\tfrac{d}{dt}\left(e^{\frac{\mu}{L}t}V(t)\right) \le 0.$$

Отже,

$$e^{\frac{\mu}{L}t}V(t) \le V(0) \quad \forall t > 0.$$

Нерівність (8) доведена. Нерівність (9) випливає з (8) та оцінки

$$\tfrac{\mu}{2}\|x - x_*\|^2 \le f(x) - f_* \le \tfrac{L}{2}\|x - x_*\|^2 \quad \forall x \in \mathbb{R}^n.$$

Лема доведена. □

Зауваження 3. Для ітераційного методу (6) із зауваження 2 можна розглянути динамічну систему

$$\begin{cases} \dot{x}(t) = -\frac{\|Fx(t)\|^2}{(F'(x(t))Fx(t), Fx(t))}Fx(t), \\ x(0) = x_0. \end{cases} \tag{10}$$

У випадку L-ліпшицевого та μ-сильно опуклого диференціовного відображення F для системи (10) мають місце тотожність та оцінка

$$\|Fx(t)\| = e^{-t}\|Fx_0\|, \;\; \|x(t) - x_*\| \le e^{-\frac{\mu}{2L}t}\|x_0 - x_*\| \quad \forall t > 0.$$

3. $O\left(\ln\frac{1}{\varepsilon}\right)$–ОЦІНКА

Сформулюємо основний результат.

Теорема 1. *Нехай функція $f : \mathbb{R}^n \to \mathbb{R}$ — двічі неперервно диференційовна, сильно опукла (з константою $\mu > 0$) та L-гладка. Тоді задача Коші* (4) *має єдиний розв'язок $t \to x(t)$ на $[0, +\infty)$, причому для довільного $\varepsilon > 0$ має місце*

$$\max\{\|\nabla f(x(t))\|, f(x(t)) - f_*, \|x(t) - x_*\|\} \le \varepsilon$$

як тільки

$$t \ge \max\left\{\ln\frac{\|\nabla f(x_0)\|}{\varepsilon}, \frac{2L}{\mu}\ln\frac{\sqrt{\frac{L}{\mu}}\,(f(x_0) - f_*)}{\varepsilon}, \frac{2L}{\mu}\ln\frac{\|x_0 - x_*\|}{\varepsilon}\right\}.$$

Доведення. Нехай $\varepsilon > 0$. З нерівності (7) випливає

$$\|\nabla f(x(t))\| \le \varepsilon$$

для $t \ge \ln \frac{\|\nabla f(x_0)\|}{\varepsilon}$. Аналогічно, з нерівностей (8), (9) випливає

$$\|x(t) - x_*\| \le \varepsilon$$

для $t \ge \frac{2L}{\mu} \ln \frac{\|x_0 - x_*\|}{\varepsilon}$, та

$$f(x(t)) - f_* \le \varepsilon$$

для $t \ge \frac{2L}{\mu} \ln \frac{\sqrt{\frac{L}{\mu}(f(x_0)-f_*)}}{\varepsilon}$, що і потрібно було довести. □

Заключні зауваження

У статті досліджено асимптотичну поведінку градієнтної системи, яка є неперервним аналогом варіанту градієнтного методу з [16] для мінімізації сильно опуклих функцій. За допомогою другого методу Ляпунова встановлено оцінки швидкості збіжності градієнтної системи до точки рівноваги.

Подібні результати можна отримати для систем вигляду:

$$\begin{cases} \dot{x}(t) = -\frac{(\nabla^2 f(x(t))\nabla f(x(t)), \nabla f(x(t)))}{\|\nabla^2 f(x(t))\nabla f(x(t))\|^2} \nabla f(x(t)), \\ x(0) = x_0; \end{cases}$$

$$\begin{cases} \dot{x}(t) = -\frac{\|\nabla f(x(t))\|^2}{\|\nabla^2 f(x(t))^* \nabla f(x(t))\|^2} \nabla^2 f(x(t))^* \nabla f(x(t)), \\ x(0) = x_0. \end{cases}$$

Дані системи є неперервними аналогами методу мінімальних нев'язок [17] (варіант α-процесу).

Робота виконана за фінансової підтримки МОН України (проєкт «Обчислювальні алгоритми і оптимізація для штучного інтелекту, медицини та оборони», 0122U002026, 2022–2024 рр.).

Література

1. Polyak B. T. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics.* 1963. Volume 3. Issue 4. P. 864–878. https://doi.org/10.1016/0041-5553(63)90382-3
2. Alber S. I., Alber Ya. I. The method of differential descent for the solution of multi-dimensional variational problems. *Dokl. Akad. Nauk SSSR.* 1966. Volume 171. Number 6. P. 1247–1250.
3. Alber S. I., Alber Ya. I. A method of differential descent for solving non-linear systems. *USSR Computational Mathematics and Mathematical Physics.* 1967. Volume 7. Issue 1. P. 15–40. https://doi.org/10.1016/0041-5553(67)90062-6
4. Goudou X., Munier J. The gradient and heavy ball with friction dynamical systems: the quasiconvex case. *Math. Program.* 2009. Vol. 116. P. 173–191. https://doi.org/10.1007/s10107-007-0109-5
5. Polyak B. T., Shcherbakov P. S. Optimisation and asymptotic stability. *International Journal of Control.* 2018. Vol. 91. No. 11. P. 2404–2410. https://doi.org/10.1080/00207179.2016.1257154

6. Polyak B. T. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics.* 1964. Volume 4. Issue 5. P. 1–17. https://doi.org/10.1016/0041-5553(64)90137-5
7. Alvarez F. On the Minimizing Property of a Second Order Dissipative System in Hilbert Spaces. *SIAM Journal on Control and Optimization.* 2000. Vol. 38. Iss. 4. P. 1102–1119. https://doi.org/10.1137/S0363012998335802
8. Alvarez F., Attouch H., Bolte J., Redont P. A second-order gradient-like dissipative dynamical system with Hessian-driven damping: Application to optimization and mechanics. *Journal de Mathematiques Pures et Appliquees.* 2002. Volume 81. Issue 8. P. 747–779. https://doi.org/10.1016/S0021-7824(01)01253-3
9. Su W., Stephen Boyd S., Candes E. J. A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. *Journal of Machine Learning Research.* 2016. Vol. 17. P. 1–43.
10. Nesterov Y. Introductory lectures on convex optimization: A basic course. Boston: Kluwer Academic Publishers, 2004. 254 p.
11. Attouch H., Fadili J. From the Ravine method to the Nesterov method and vice versa: a dynamical system perspective. arXiv:2201.11643. 2022.
12. Alber Y. I. Continuous regularization of linear operator equations in a Hilbert space. *Mathematical Notes of the Academy of Sciences of the USSR.* 1968. 4. P. 793–797. https://doi.org/10.1007/BF01111311
13. Alber Ya. I. Continuous processes of Newton type. *Differ. Uravn.* 1971. Volume 7. Number 11. P. 1931–1945.
14. Suh J. J., Roh G., Ryu E. K. Continuous-Time Analysis of Accelerated Gradient Methods via Conservation Laws in Dilated Coordinate Systems. *Proceedings of the 39th International Conference on Machine Learning.* 2022. P. 20640–20667.
15. Attouch H., Chbani Z., Fadili J., Riahi H. Fast Convergence of Dynamical ADMM via Time Scaling of Damped Inertial Dynamics. *J. Optim. Theory Appl.* 2022. Vol. 193. P. 704–736. https://doi.org/10.1007/s10957-021-01859-2
16. Fatkhullin I., Polyak B. Optimizing Static Linear Feedback: Gradient Method. arXiv: 2004.09875. 2020.
17. Krasnoselskii M. A., Vainikko G. M., Zabreiko P. P., Rutitcki Ja. B., Stecenko V. Ja. Approximated Solutions of Operator Equations. Groningen: Walters–Noordhoff, 1972. 484 p.

Надійшла: 23.02.2022 / Прийнята: 27.04.2022