

УДК 519.85

MSC 47J20, 49J40, 65K15, 90C25

Proximal Algorithms for bi-level Convex Optimization Problems

A.V. Luita, S.O. Zhilina, V.V. Semenov

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

E-mail: {nastialuita, zhilina.1958, semenov.volodya}@gmail.com

In this paper, problems of bi-level convex minimization in a Hilbert space are considered. The bi-level convex minimization problem is to minimize the first convex function on the set of minima of the second convex function. This setting has many applications, but the implicit constraints generated by the internal problem make it difficult to obtain optimality conditions and construct algorithms. Multilevel optimization problems are formulated in a similar way, the source of which is the operation research problems (optimization according to sequentially specified criteria or lexicographic optimization). Attention is focused on problem solving using two proximal methods. The main theoretical results are theorems on the convergence of methods in various situations. The first of the methods is obtained by combining the penalty function method and the proximal method. Strong convergence is proved in the case of strong convexity of the function of the exterior problem. In the general case, only weak convergence has been proved. The second, the so-called proximal-gradient method, is a combination of one of the variants of the fast proximal-gradient algorithm with the method of penalty functions. The rates of convergence of the proximal-gradient method and its weak convergence are proved.

Key words: convex optimization, bi-level problem, proximal algorithm, convergence.

Проксимальні алгоритми для дворівневих задач опуклої оптимізації

А.В. Люта, С.О. Жиліна, В.В. Семенов

Київський національний університет імені Тараса Шевченка, Київ, Україна

E-mail: {nastialuita, zhilina.1958, semenov.volodya}@gmail.com

У роботі розглянуто задачі дворівневої опуклої мінімізації у гільбертовому просторі. Дворівнева задача опуклої мінімізації полягає у мінімізації першої опуклої функції на множині мінімумів другої опуклої функції. Ця постановка має багато застосувань, але неявні обмеження, що породжені внутрішньою задачею ускладнюють отримання умов оптимальності та побудову методів. Подібним чином формулюються й багаторівневі задачі, джерелом яких стали питання дослідження операцій (оптимізація за послідовно заданими критеріями або лексикографічна оптимізація). Увага зосереджена на розв'язанні задач за допомогою двох методів проксимального типу. Основні теоретичні результати – теореми про збіжність методів у різних ситуаціях. Перший з методів отриманий поєднанням методу штрафних функцій та проксимального методу. Доведена сильна збіжність у випадку сильної опуклості функції зовнішньої задачі. У загальному випадку отримана лише слабка збіжність. Другий, так званий, проксимально-градієнтний метод є поєднанням одного з варіантів швидкого проксимально-градієнтного алгоритму з методом штрафних функцій. Встановлені оцінки швидкості проксимально-градієнтного методу та його слабка збіжність.

Ключові слова: опукла оптимізація, дворівнева задача, проксимальний алгоритм, збіжність.

1. Вступ

У роботі розглянуто питання розв'язання дворівневої опуклої задачі мінімізації у гільбертовому просторі. Дворівнева задача опуклої оптимізації полягає у мінімізації першої опуклої функції на множині мінімумів другої опуклої функції. Ця постановка має багато застосувань, але неявні обмеження, що породжені внутрішньою задачею ускладнюють отримання умов оптимальності та побудову методів. Подібним чином формулюються й багаторівневі задачі, джерелом яких стали питання дослідження операцій (оптимізація за послідовно заданими критеріями або лексикографічна оптимізація) [1]. Увага зосереджена на розв'язанні задач за допомогою двох методів проксимального типу. Основні теоретичні результати – теореми про збіжність методів у різних ситуаціях.

2. Постановка задач

В оптимізації та теорії некоректних задач є популярним такий підхід до розв'язання задач з неєдиним розв'язком. Розглядають певну родину збурених задач, однозначно та коректно розв'язних. Частинний розв'язок вихідної задачі одержують як границю розв'язків збурених задач при зменшенні збурень. Знайдені так частинні розв'язки задовольняють певним додатковим умовам, наприклад, мінімальність норми нормального розв'язку задачі [2, 3]. Іншим джерелом задач вигляду

$$f_2(x) \rightarrow \min, \quad x \in \arg \min f_1$$

є метод штрафів та задачі оптимізації за послідовно заданими критеріями (лексикографічна, послідовна або багаторівнева оптимізація) [1].

У даній роботі розглядається дворівнева задача опуклої оптимізації в гільбертовому просторі. Ефективні методи для таких задач у скінченновимірній постановці запропоновані в [4, 5].

Нехай H – дійсний гільбертовий простір зі скалярним добутком (\cdot, \cdot) та нормою $\|\cdot\|$. Нехай $f_1, f_2 : H \rightarrow \mathbb{R} \cup \{+\infty\}$ – власні опуклі напівнеперервні знизу функціонали. Припустимо, що $\arg \min f_1 \neq \emptyset$ та $\min f_1 = 0$. Розглянемо задачу

$$f_2(x) \rightarrow \min, \quad x \in \arg \min f_1. \tag{1}$$

Припустимо, що $0 \in \text{int}(\text{dom } f_2 - \arg \min f_1)$. Тоді задача (1) еквівалентна включенню

$$\text{знайти } x \in H : \quad 0 \in \partial f_2(x) + N_{\arg \min f_1} x,$$

де $N_M x$ – нормальний конус замкненої опуклої множини $M \subseteq H$ в точці $x \in H$, тобто

$$N_M x = \begin{cases} \{z \in H : (z, y - x) \leq 0 \quad \forall y \in M\}, & \text{якщо } x \in M, \\ \emptyset, & \text{інакше.} \end{cases}$$

Позначимо через S множину $\arg \min f_1$, а через C множину розв'язків (1).

Нехай $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ – власний опуклий напівнеперервний знизу функціонал. Проксимальним оператором, асоційованим з функціоналом g , називають оператор $H \ni x \mapsto \text{prox}_g x = \arg \min_{y \in H} \left(g(y) + \frac{1}{2} \|y - x\|^2 \right)$.

3. Проксимальний алгоритм

Метою є дослідження збіжності схем вигляду

$$x_{n+1} = \arg \min_{y \in H} \left\{ \lambda_n f_2(y) + \lambda_n \alpha_n f_1(y) + \frac{1}{2} \|y - x_n\|^2 \right\}.$$

При $\alpha_n = 0$ маємо звичайний проксимальний метод для задачі $f_2 \rightarrow \min$ [6].

Нехай (λ_n) , (α_n) – послідовності додатніх чисел.

Алгоритм 1. *Задаємо $x_1 \in H$.*

Ітераційний крок. *Для $x_n \in H$ обчислюємо*

$$x_{n+1} = \text{PROX}_{\lambda_n}(f_2 + \lambda_n \alpha_n f_1) x_n.$$

У випадку сильної опуклості f_2 алгоритм 1 сильно збігається.

Теорема 1. *Нехай функціонал f_2 сильно опуклий. Припустимо, що $\sum_{n=1}^{\infty} \lambda_n = +\infty$,*

$\lim_{n \rightarrow \infty} \lambda_n = 0$, $\lim_{n \rightarrow \infty} \alpha_n = +\infty$, $\liminf_{n \rightarrow \infty} \alpha_n \lambda_n > 0$. Тоді породжена алгоритмом 1 послідовність (x_n) сильно збігається до єдиного розв'язку задачі (1).

Перейдемо до ситуації, коли функціонал f_2 не є сильно опуклим. Припустимо, що

$$(A1) \exists k > 0: f_1(x) \geq k d_C^2(x) = k \min_{y \in C} \|x - y\|^2 \quad \forall x \in H.$$

Лема 1. *Нехай для f_1 виконується припущення (A1). Тоді для $z \in C$ і $w \in N_C z$ має місце нерівність*

$$(w, x) - f_1(x) - (w, z) \leq (4k)^{-1} \|w\|^2 \quad \forall x \in H.$$

Має місце

Лема 2. *Нехай для f_1 виконується (A1). Нехай $z \in C$, $v \in \partial f_2(x) + N_C x$, а точка $w \in N_C z$, така, що $v - w \in \partial f_2(x)$. Тоді виконується нерівність*

$$\|x_{n+1} - z\|^2 - \|x_n - z\|^2 + \|x_{n+1} - x_n\|^2 + \lambda_n \alpha_n f_1(x_{n+1}) \leq 2\lambda_n (v, z - x_{n+1}) + \frac{\lambda_n}{\alpha_n} \frac{1}{k} \|w\|^2.$$

Теорема 2. *Нехай $S \neq \emptyset$, виконується умова (A1). Припустимо, що $\sum_{n=1}^{\infty} \lambda_n \alpha_n^{-1} < +\infty$, $\liminf_{n \rightarrow \infty} \lambda_n > 0$. Тоді породжена алгоритмом 1 послідовність (x_n) слабо збігається до розв'язку задачі (1).*

Використовуючи техніку, розвинену в [7], можна показати, що розв'язність задачі (1) рівносильна слабкій збіжності послідовності середніх за Чезаро елементів x_n , породжених алгоритмом 1.

Лема 3. *Для точок z , v , w з лем 2 та породженої алгоритмом 1 послідовності (x_n) виконується нерівність*

$$-\left(2 \sum_{i=1}^n \lambda_i\right)^{-1} \|x_1 - z\|^2 \leq (v, z - \bar{x}_{n+1}) + \left(\sum_{i=1}^n \lambda_i\right)^{-1} \left(\sum_{i=1}^n \lambda_i \alpha_i^{-1}\right) \frac{1}{2k} \|w\|^2,$$

$$\text{де } \bar{x}_n = \left(\sum_{k=1}^{n-1} \lambda_k \right)^{-1} \sum_{k=1}^{n-1} \lambda_k x_{k+1}.$$

Теорема 3. Нехай виконується умова (A1). Припустимо, що $\sum_{n=1}^{\infty} \lambda_n \alpha_n^{-1} < +\infty$, $\sum_{n=1}^{\infty} \lambda_n = +\infty$. Тоді справедливі твердження: якщо $S \neq \emptyset$, то послідовність (\bar{x}_n) слабо збігається до точки з S ; якщо $S = \emptyset$, то $\|\bar{x}_n\| \rightarrow +\infty$.

4. Проксимально-градієнтний алгоритм

Нехай H – дійсний гільбертовий простір. Розглянемо задачу:

$$h_2(x) = f_2(x) + g_2(x) \rightarrow \min, \quad (2)$$

$$x \in \arg \min h_1 = \arg \min (f_1 + g_1), \quad (3)$$

де $g_1, g_2: H \rightarrow \mathbb{R} \cup \{+\infty\}$ – власні опуклі напівнеперервні знизу функції, $f_1, f_2: H \rightarrow \mathbb{R}$ – опуклі диференційовні функції з ліпшицевими градієнтами. Позначимо через L_1, L_2 константи Ліпшиця градієнтів $\nabla f_1, \nabla f_2$.

Для розв'язання дворівневої задачі мінімізації (2), (3) пропонуємо такий ітераційний проксимально-градієнтний алгоритм [8–10].

Алгоритм 2. Задаємо $\alpha_1 \in (0, 1]$, $\lambda > 0$, $\mu > 0$, елементи $x_1 = y_1 \in H$.

Ітераційний крок. Обчислити

$$z_n = \alpha_n x_n + (1 - \alpha_n) y_n,$$

$$x_{n+1} = \text{прох}_{\left(\lambda g_2 + \frac{\mu}{\alpha_n} g_1\right)} \left(x_n - \left(\lambda \nabla f_2(x_n) + \frac{\mu}{\alpha_n} \nabla f_1(z_n) \right) \right),$$

$$y_{n+1} = \alpha_n x_{n+1} + (1 - \alpha_n) y_n,$$

$$\alpha_{n+1} = \frac{\sqrt{\alpha_n^4 + 4\alpha_n^2} - \alpha_n^2}{2}.$$

Параметри λ та μ можна обирати за правилом:

$$\lambda L_2 + \mu L_1 \leq 1. \quad (4)$$

Наприклад, $\lambda = 1/kL_2$, $\mu = 1/lL_1$, де $k, l \in \{2, 3, \dots\}$.

Алгоритм 2 є результатом адаптації для дворівневих задач однієї з прямо-двоїстих проксимально-градієнтних схем Tseng'а для розв'язання опуклих задач вигляду $\min(f + g)$ [11]. А саме такої:

$$z_n = \frac{2}{n+2} x_n + \frac{n}{n+2} y_n,$$

$$x_{n+1} = \text{прох}_{\frac{n+2}{2L} g} \left(x_n - \frac{n+2}{2L} \nabla f(z_n) \right),$$

$$y_{n+1} = \frac{2}{n+2} x_{n+1} + \frac{n}{n+2} y_n,$$

де L – константа Ліпшиця ∇f .

Позначимо S_1 множину розв'язків внутрішньої задачі, тобто $\arg \min h_1$, а S_2 множину розв'язків зовнішньої задачі. Відносно збіжності алгоритму 2 з правилом (4) відомо таке. Якщо $S_2 \neq \emptyset$, то

$$f_1(y_n) + g_1(y_n) - \min(f_1 + g_1) = O\left(\frac{1}{n}\right) \text{ при } n \rightarrow \infty.$$

Якщо додатково множина S_1 обмежена, то

$$f_1(y_n) + g_1(y_n) - \min(f_1 + g_1) = o\left(\frac{1}{n}\right) \text{ при } n \rightarrow \infty,$$

$$f_2(y_n) + g_2(y_n) \rightarrow \min_{S_1}(f_2 + g_2),$$

всі послідовність (y_n) обмежена та всі її часткові слабкі границі належать S_2 .

5. Висновки

У роботі для загальної дворівневої опуклої задачі мінімізації у гільбертовому просторі досліджено два методи проксимального типу. Перший з методів отриманий поєднанням методу штрафних функцій та проксимального методу. Доведена сильна збіжність у випадку сильної опуклості функції зовнішньої задачі. У загальному випадку отримана лише слабка збіжність. Другий, так званий, проксимально-градієнтний метод є поєднанням одного з варіантів швидкого проксимально-градієнтного алгоритму з методом штрафних функцій. Попередні результати свідчать про перспективність проксимально-градієнтного методу (алгоритм 2). Якщо константи Ліпшиця градієнтів ∇f_1 , ∇f_2 не відомі, то виникає питання підбору параметрів λ та μ . Цікавим є варіант використання адаптивних правил (як у відомих методах Adagrad, Adam).

Робота виконана при фінансовій підтримці МОН України («Математичне моделювання та оптимізація динамічних систем для оборони, медицини та екології», 0119U100337).

ЛІТЕРАТУРА

1. Podinovskii V.V., Gavrilov V.M. Optimization with respect to successively applied criteria. Moscow: Sovetskoe Radio, 1975. 192 p.
2. Bakushinskii A. B., Goncharkii A. V. Iterative Methods for Solving Ill-Posed Problems. Moscow: Nauka, 1989. 126 p.
3. Attouch H. Viscosity Solutions of Minimization Problems. *SIAM J. on Optim.* 1996. Vol. 6. P. 769–806.
4. Solodov M. An explicit descent method for bilevel convex optimization. *Journal of Convex Analysis.* 2007. Vol. 4. P. 227–238.
5. Solodov M. A bundle method for a class of bilevel nonsmooth convex minimization problems. *SIAM J. on Optim.* 2007. Vol. 18. P. 242–259.
6. Guler O. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.* 1991. Vol. 29. P. 403–419.
7. Passty G.B. Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Spaces. *Journal of Mathematical Analysis and Applications.* 1979. Vol. 72. P. 383–390.

8. Malitsky Yu. Chambolle-Pock and Tseng's methods: relationship and extension to the bilevel optimization. *arXiv:1706.02602*. 2017.
9. Luita A.V., Semenov V.V. A novel method for the bilevel optimization problem. XXXII International Conference “Problems of decision making under uncertainties”. August 27-31, 2018. Prague, Czech Republic. Abstracts. P. 81–82.
10. Luita A.V., Semenov V.V. A novel proximal-gradient method for the bilevel optimization problem. Proceedings of the XXIV All-Ukrainian Scientific Conference “Contemporary Problems of Applied Mathematics and Informatics”. September 26-28, 2018. Lviv. P. 77–79.
11. Tseng P. On accelerated proximal gradient methods for convex-concave optimization. 2008. <http://www.mit.edu/~dimitrib/PTseng/papers/apgm.pdf>.

Надійшла 21.05.2021.

Проксимальные алгоритмы для двухуровневых задач выпуклой оптимизации

А.В. Лютая, С.А. Жилина, В.В. Семенов

Киевский национальный университет имени Тараса Шевченко, Киев, Украина

E-mail: {nastialuita, zhilina.1958, semenov.volodya}@gmail.com

В данной работе рассмотрены задачи двухуровневой выпуклой минимизации в гильбертовом пространстве. Двухуровневая задача выпуклой минимизации заключается в минимизации первой выпуклой функции на множестве минимумов второй выпуклой функции. Эта постановка имеет много применений, но неявные ограничения, порожденные внутренней задачей, затрудняют получение условий оптимальности и построение алгоритмов. Подобным образом формулируются и многоуровневые задачи оптимизации, источником которых стали вопросы исследования операций (оптимизация по последовательно заданным критериям или лексикографическая оптимизация). Внимание сосредоточено на решении задач с помощью двух методов проксимального типа. Основные теоретические результаты – теоремы о сходимости методов в различных ситуациях. Первый из методов получен сочетанием метода штрафных функций и проксимального метода. Доказана сильная сходимость в случае сильной выпуклости функции внешней задачи. В общем случае установлена лишь слабая сходимость. Второй, так называемый, проксимально-градиентный метод является сочетанием одного из вариантов быстрого проксимально-градиентного алгоритма с методом штрафных функций. Установлены оценки скорости проксимально-градиентного метода и его слабая сходимость.

Ключевые слова: выпуклая оптимизация, двухуровневая задача, проксимальный алгоритм, сходимость.