УДК 518.90; 519.95

П.О. Кравець
НУ "Львівська політехніка",
кафедра "Інформаційні системи та мережі"

ЗБІЖНІСТЬ ІГРОВОГО ГРАДІЄНТНОГО МЕТОДУ У ЗНАКОДОДАТНИХ СЕРЕДОВИЩАХ

This article investigates convergence conditions of a gradient game method in conditions of uncertainty of positive environments. The method is constructed because of dynamic mixed strategies of the players, which realize selection of optimum line solutions. The convergence conditions of a game method with probability 1 and in average quadratic to a point set of equilibrium on Nash are obtained. The theoretical researches are confirmed by results of computer simulation.

У статті досліджується проблема збіжності градієнтного ігрового методу у знакододатних середовищах в умовах їх апріорної невизначеності. Метод побудовано на основі динамічних змішаних стратегій гравців, які здійснюють селекцію оптимальних поточних розв'язків у процесі самонавчання. Отримано умови збіжності ігрового методу з імовірністю 1 та у середньоквадратичному до множини точок рівноваги за Нешем. Теоретичні результати апробовано програмним моделюванням на комп'ютері.

Вступ. Для оптимального проектування систем різного призначення є необхідним розв'язування задачі багатокритеріальної оптимізації в умовах невизначеності, обумовленої апріорною недостатністю специфікації параметрів або критеріїв системи [1]. Для цього використовуються різноманітні пошукові методи, серед яких найбільш ефективними є ті, котрі шляхом опрацювання даних експерименту здійснюють селекцію кращих поточних розв'язків. Такі методи є адаптивними до невизначеностей системи та забезпечують зменшення кількості пошукових кроків. Якщо аргументи критеріальних функцій набувають скінченну кількість дискретних значень, то для розв'язування задачі пошукової оптимізації ефективними будуть адаптивні ігрові методи.

Формулювання ігрової задачі в умовах невизначеності. Нехай задана множина гравців D, які взаємодіють між собою через кероване стохастичне середовище S. Динаміка середовища S гравцям апріорі не відома. Гравці можуть впливати на середовище за допомогою реалізації однієї із власних чистих стратегій $U^i = (u^i(1), u^i(2), ..., u^i(N_i))$, які розглядаються як деякі керуючі дії. Вибір чистих стратегій здійснюється гравцями незалежно у моменти часу $n = 1,2,...$ за допомогою випадкового механізму, побудованого на основі імовірнісного розподілу $p_n^i = (p_n^i(1), p_n^i(2), ..., p_n^i(N_i))$. Вектори $p_n^i \in S_\varepsilon^{N_i}$ є змішаними стратегіями гравців і набувають значення на одиничних ε-симплексах:

$$S_\varepsilon^{N_i} = \left\{p^i \mid p^i \in S^{N_i}; p^i(j) \ge \varepsilon \ (j = \overline{1, N_i}\right\}, \ \varepsilon \in (0, \min_i N_i^{-1}). \qquad (1)$$

Виграші i-го гравця визначаються на декартовому добутку $U^{D_i} = \underset{j \in Di}{\otimes} U^j$ стратегій локальної підмножини гравців $D_i \subseteq D$. Після реалізації спільного вибору $u_n^{D_i} \in U^{D_i}$ кожен i-й гравець отримує випадковий поточний виграш $\xi_n^i = \xi_n^i(u_n^{D_i})$, який використовується для перерахунку векторів змішаних стратегій за загальним правилом:

елемент $p_n^i(u_n^i)$ зростає пропорційно величині виграшу ξ_n^i, а інші елементи зменшуються для забезпечення умови $p_n^i \in S_\varepsilon^{N_i}$. Така зміна елементів векторів змішаних стратегій призводить до більш частого вибору тих чистих стратегій, які дають найбільший середній виграш

$$\Phi_n^i(\{u_n^{D_i}\}) = \frac{1}{n}\sum_{t=1}^{n}\xi_t^i \quad \forall i \in D. \qquad (2)$$

Теоретично, гра в умовах невизначеності розгортається на безмежному відрізку часу, а поведінка гравців спрямована на досягнення асимптотично максимального середнього виграшу

$$\underline{\lim}_{n\to\infty} \Phi_n^i \to \max \forall i \in D.$$

На практиці час гри обмежується умовами точності досягнутого розв'язку.

Враховуючи багатокритеріальний характер ігрової задачі, її глобальний розв'язок будемо шукати у множині точок асимптотичної рівноваги за Нешем

$$\forall i \in D \qquad \underline{\lim}_{n\to\infty}\left[\Phi_n^i(\{u_n^{D_i}\}) - \Phi_n^i(\{\hat{u}_n^{D_i}\})\right] \ge 0,$$

де $\hat{u}_n^{D_i} = u_n^{D_i} \setminus u_n^i + \tilde{u}_n^i \in U^{D_i}$ — поточне значення спільної стратегії підмножини гравців D_i, отримане при заміні чистої стратегії i-го гравця з u_n^i на $\tilde{u}_n^i \in U^i$.

Надалі будемо вважати, що випадкові виграші $\{\xi_n^i\}$ є незалежними $\forall u_n^{D_i} \in U^{D_i}$, $\forall i \in D$, $n = 1,2,\ldots$, мають постійне математичне сподівання $M\{\xi_n^i(u^{D_i})\} = v(u^{D_i}) = const$ та обмежений другий момент $\sup_n M\{[\xi_n^i(u^{D_i})]^2\} = \sigma_i^2(u^{D_i}) < \infty$. Закони розподілу випадкових величин та значення їх моментів є апріорі невідомими.

З урахуванням цього середовище S задається сукупністю випадкових розподілів $Z(v(u^{D_i}), d(u^{D_i}))$ з математичними сподіваннями $v(u^{D_i})$ та дисперсіями $d(u^{D_i})$.

Побудова методу розв'язування ігрової задачі. Для знаходження рівноважних за Нешем розв'язків необхідно задати спосіб зміни векторів змішаних стратегій у часі. Для цього використаємо рекурентний метод, побудований на основі градієнта функції середніх виграшів, визначеної для матричної гри в умовах повної інформації:

$$V^i(p^{D_i}) = \sum_{u^{D_i} \in U^{D_i}} v^i(u^{D_i}) \prod_{j \in D_i; u^j \in u^{D_i}} p^j(u^j),$$

де $p^{D_i} \in S_\varepsilon^{D_i}$, $S_\varepsilon^{D_i} = \prod_{j \in D_i} S_\varepsilon^{N_j}$, $u^{D_i} \in U^{D_i}$.

Для рівноважних за Нешем розв'язків у повністю змішаних стратегіях виконується умова доповняльної нежорсткості:

$$\nabla_{p^i} V^i = V^i e^{N_i}, p^i \in S^{N_i}, \forall i \in D, \qquad (3)$$

де $\nabla_{p^i} V_n^i$ — градієнт функції середніх виграшів; $e^{N_i} = (1_j \mid j = \overline{1, N_i})$ — вектор, всі елементи якого дорівнюють 1.

Виконаємо нормалізацію векторної умови (3), домноживши її поелементно на вектор p^i:

$$\operatorname{diag}(p^i)(e^{N_i} V^i(p^{D_i}) - \nabla_{p^i} V^i(p^{D_i})) = 0, \qquad (4)$$

де $\operatorname{diag}(p^i)$ — діагональна матриця порядку N_i, складена з елементів вектора $p^i \in S^{N_i}$.

Нормалізація (4) забезпечує належність вектора змішаних стратегій p^i одиничному симплексу S^{N_i} та дає можливість врахувати можливі розв'язки на його межі. Означення одиничного симплексу S^{N_i} отримують з (1) при $\varepsilon = 0$.

Враховуючи, що $\nabla_{p^i} V^i = M\left\{ \dfrac{\xi_n^i}{e^T(u_n^i)p_n^i} e(u_n^i) \middle| p_n^i = p^i \right\}$, на основі методу стохастичної апроксимації [2] отримаємо такий рекурентний метод:

$$p_{n+1}^i = \pi_{\varepsilon_{n+1}}^{N_i} \left\{ p_n^i + \gamma_n \frac{\xi_n^i e(u_n^i)}{e^T(u_n^i)p_n^i} \right\}, \qquad (5)$$

де $p^i \in S_\varepsilon^{N_i}$, $e(u_n^i)$ — одиничний вектор-індикатор події $u_n^i = u^i$, $\pi_{\varepsilon_{n+1}}^{N_i}$ — проектор на одиничний ε-симплекс, γ_n — монотонно спадна послідовність невід'ємних величин, яка регулює величину кроку методу; $\varepsilon_n \in (0, \min_i N_i^{-1})$ — монотонно спадна послідовність величин, що регулює швидкість розширення одиничного ε-симплексу. Проектування вектора $q \in R^{N_i}$ на одиничний ε-симплекс полягає у розв'язуванні задачі квадратичного програмування

$$\left\| q_n^i - p_n^i \right\|^2 \to \min_{p_n^i}$$

з обмеженнями $\sum_{j=1}^{N_i} p_n^i(j) = 1$, $p_n^i(j) \ge \varepsilon_n$.

Параметри методу γ_n та ε_n визначатимуть умови та швидкість збіжності ігрового методу. Приймемо, що ці параметри є монотонними послідовностями виду

$$\gamma_n = \gamma(n+a)^{-\alpha};\ a > 0;\ \varepsilon_n = \varepsilon(n+b)^{-\beta};\ b > 0. \qquad (6)$$

Збіжність методу (5) досліджувалася у [3] для середовищ загального виду. Було доведено, що у середовищах загального виду максимальний степеневий порядок швидкості збіжності методу (5) дорівнює $\theta = 1/3$, що досягається при $\alpha = 2/3$, $\beta = 1/3$.

Якщо середовище є організоване певним чином, наприклад, матриці середніх виграшів є діагональноопуклими, то ігровий метод (5) може мати більшу швидкість збіжності. Для ряду практичних задач пошукової оптимізації значення виграшів має один і той же знак для всіх можливих комбінацій варіантів чистих стратегій. Середовища, які породжують такі виграші, будемо називати знакопостійними.

Визначимо умови збіжності методу (5) у знакододатних середовищах, для яких $v_{min} = \min_{i \in D} \min_{u^{D_i} \in U^{D_i}} v^i(u^{D_i}) > 0$. Збіжність у знаковід'ємних середовищах при таких же умовах забезпечується зміною знаку вектора руху методу (5) на протилежний.

Умови збіжності ігрового методу (5) визначаються наступними твердженнями.

Твердження 1. Для незалежних випадкових величин поточних виграшів ξ_n^i, незалежного вибору чистих стратегій u_n^i проекційний ігровий метод (5) при $n \to \infty$; $\gamma_n > 0$; $\gamma_{n+1} < \gamma_n$; $\sum_{n \to \infty} \gamma_n = \infty$, $\varepsilon_n \in (0, \min_{i \in D} N_i^{-1})$; $\varepsilon_{n+1} < \varepsilon_n$; $\lim_{n \to \infty} \varepsilon_n = 0$ для будь-якого початкового наближення $p_1^i \in S_{\varepsilon_1}^{N_i}$ $\forall i \in D$ забезпечує виконання умови доповняльної нежорсткості (4) у знакододатному середовищі $v_{min} > 0$:

з ймовірністю 1, якщо $\sum_{n=1}^{\infty} (\gamma_n^2 \varepsilon_n^{-1} + |\varepsilon_n - \varepsilon_{n+1}|) < \infty$;

у середньоквадратичному, якщо $\lim_{n \to \infty} (\gamma_n \varepsilon_n^{-1} + |\varepsilon_n - \varepsilon_{n+1}| \gamma_n^{-1}) = 0$.

Доведення твердження 1 базується на основі зваженої умови доповняльної нежорсткості (4). Визначимо функцію Ляпунова $\Delta_n = \sum_{i \in D} \left\| p_n^i - \tilde{p}_n^i \right\|^2$, у якій вектори $p_n^i \in S_{\varepsilon_n}^{N_i}$ генеруються методом (5), а $\tilde{p}_n^i \in S^{N_i}$ визначаються з умови (4)

$$\tilde{p}_n^i(j) = V_{nj}^i \big/ V_n^i, \quad j = \overline{1, N_i},$$

де $V_{nj}^i = p_n^i(j) \nabla V_{nj}^i$.

Введемо σ-алгебру $F_n = \left\{ u_t^i, \xi_t^i \mid t = \overline{1, n-1}; \forall i \in D \right\}$ та знайдемо верхню оцінку умовного математичного сподівання $M\left\{ \Delta_{n+1} \middle| F_n \right\}$:

$$M\left\{ \Delta_{n+1} \middle| F_n \right\} \le (1 - 2\gamma_n v_{min}) \Delta_n + C(\mu_n + \gamma_n^2 \varepsilon_n^{-1}), \tag{7}$$

де $C \sim |D| N_{max} v_{max} \sigma_{max}^2 > 0$, $\mu_n = |\gamma_{n-1} - \gamma_n| + |\varepsilon_n - \varepsilon_{n+1}|$.

Усереднюючи (7) по F_n, отримаємо:

$$M\{\Delta_{n+1}\} \le (1 - 2\gamma_n v_{min}) M\{\Delta_n\} + C(\mu_n + \gamma_n^2 \varepsilon_n^{-1}). \tag{8}$$

Якщо $v_{min} > 0$ то збіжність методу (5) з ймовірністю 1 випливає з умов твердження 1, оцінки (7) та наслідків теореми Роббінса-Сігмунда (лема П.11 [3]) .

Збіжність у середньоквадратичному отримується застосуванням до оцінки (8) умов теореми про рекурентні числові нерівності (лема П.5 [3]).

У класі послідовностей (6) збіжність методу (5) має місце:

1) з ймовірністю 1, якщо $\alpha \in (0.5;1]$; $0 < \beta < 2\alpha - 1$;

у середньоквадратичному, якщо $\alpha \in (0,1]$; $0<\beta<\alpha$.

З урахуванням (6) оцінку (8) запишемо у вигляді:

$$M\{\Delta_{n+1}\} \le \left(1-2w_{\min}\frac{\gamma}{n^{\alpha}}\right)M\{\Delta_n\}+\frac{C(m)}{n^{\rho}}, \qquad (9)$$

де $m=\varepsilon\beta\chi(\alpha-\beta\le 0.5)+\gamma^2\varepsilon^{-1}\chi(\alpha-\beta\ge 0.5)$, $\rho=\min(1+\beta,2\alpha-\beta)$.

Твердження 2. Нехай виконані умови збіжності методу (5) у середньоквадратичному для послідовностей $\{\gamma_n\}$, $\{\varepsilon_n\}$ виду (6), зокрема $\alpha \in (0,1]$, $0<\beta<\alpha$, і при $\alpha=1$ виконується нерівність $2\gamma v_{\min} > 1-\min(\beta,1-\beta)$, тоді має місце оцінка:

$$\overline{\lim_{n\to\infty}}\, n^{\theta}\Delta_n \le \frac{C(m)}{2\gamma v_{\min}-\min(\beta,1-\beta)\chi(\alpha=1)}<\infty, \qquad (10)$$

де $\theta=\min(1+\beta-\alpha,\alpha-\beta)$ — порядок швидкості збіжності.

Доведення твердження 2 базується на застосуванні до оцінки (9) леми П.2 [3] про рекурентні числові нерівності.

Наслідок. В умовах твердження 2 порядок швидкості збіжності задовольняє умову: $0<\theta\le 0.5$. Максимальний порядок збіжності методу (5) досягається при $\alpha-\beta=0.5$ і становить $\theta=0.5$.

Результати комп'ютерного моделювання. *Початкові умови ігрового методу* визначаються кількістю гравців, кількістю стратегій, параметрами середовища гри, параметрами методу, значеннями елементів векторів змішаних стратегій.

Результати моделювання отримані для гри $|D|=5$ гравців, кожен з яких має по $N_i=2$ чисті стратегії.

Знакододатне середовище гри визначається нормальним законом розподілу виграшів з математичними сподіваннями $v^i(u^{D_i})\in[0.1,0.9]$ та дисперсією $d^i(u^{D_i})=1$.

Початкові значення параметрів $\gamma_0,\varepsilon_0,\alpha,\beta$ рекурентного методу (5) вибираються такими, щоб задовольнити умови його збіжності у середньоквадратичному.

Початкові значення елементів векторів змішаних стратегій приймаються однаковими:

$$p_n^i(j)=1/N_i,\ j=\overline{1,N_i},\ \forall i\in D,$$

що моделює ситуацію невизначеності у початковий момент часу, коли гравці не мають інформації про середовище прийняття рішень і будь-який з допустимих варіантів рішення є прийнятним для реалізації.

Визначення параметрів швидкості збіжності ігрового методу. Збіжність гри визначається поведінкою у часі випадкових процесів $\Delta_n=\sum_{i\in D}\left\|p_n^i-\tilde{p}_n^i\right\|^2$ та $\overline{\Delta}_n=\frac{1}{n}\sum_{t=1}^{n}\Delta_t$, де $\tilde{p}_n^i=\mathrm{diag}(p_n^i)\nabla_{p^i}V_n^i\Big/V_n^i$. Середній поточний виграш, який припадає на одного гравця, визначається так: $\Phi_n=|D|^{-1}\sum_{i\in D}\Phi_n^i$, де $|D|$ — потужність множини гравців.

Довжина досліджуваної вибірки становить 10 тис. кроків. На основі оцінки швидкості збіжності (10) поведінка процесу Δ_n у часі апроксимована залежністю $\Delta_n = \vartheta / n^{\theta}$, де $\vartheta > 0; \theta \in (0,1]; n = 1,2,\ldots$. Після логарифмування отримаємо лінійне співвідношення

$$\lg \Delta_n = \lg \vartheta - \theta \lg n . \qquad (11)$$

З урахуванням цього побудуємо залежність $\lg \Delta_n = f(\lg n)$. Тоді параметр $\theta = \frac{\lg \Delta_n}{\lg n}$ вказуватиме на порядок швидкості збіжності досліджуваного методу.

Для визначення порядку швидкості збіжності θ виконаємо апроксимацію випадкового процесу $\lg \Delta_n$ лінійною залежністю (11) на відрізку $\lg n \in [3,4]$ з кроком $\delta = 0.1$ за методом найменших квадратів:

$$\theta = \frac{(1+\delta^{-1}) \sum\limits_{\lg n = \overline{3,4},\delta} \lg \Delta_n \lg n - \sum\limits_{\lg n = \overline{3,4},\delta} \lg \Delta_n \sum\limits_{\lg n = \overline{3,4},\delta} \lg n}{\left(\sum\limits_{\lg n = \overline{3,4},\delta} \lg n \right)^2 - (1+\delta^{-1}) \sum\limits_{\lg n = \overline{3,4},\delta} (\lg n)^2} . \qquad (12)$$

Для згладжування випадкової складової швидкості збіжності та виділення порядку цієї швидкості виконаємо усереднення випадкового процесу Δ_n по $m = 100$ реалізаціях:

$$\overline{\overline{\Delta}}_n = \frac{1}{m} \sum_{j=1}^{m} \Delta_{n,j} .$$

Моделювання оптимального методу. На рис. 1 зображені графіки, що демонструють поведінку оптимального градієнтного методу для таких значень параметрів: $\gamma_0 = 5$, $\alpha = 1$, $\varepsilon_0 = 0.49$, $\beta = 0.5$. Експериментальний порядок швидкості збіжності методу (5), визначений згідно із (12), становить $\theta \approx 0.6$, що відповідає теоретичним результатам.

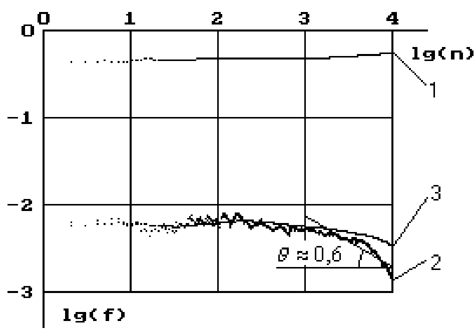

Рис. 1. *Поведінка оптимального градієнтного методу (5): 1 - функція середніх виграшів Φ_n; 2 - поточна похибка доповняльної нежорсткості Δ_n; 3 - усереднена у часі похибка доповняльної нежорсткості $\overline{\Delta}_n$.*

Вплив параметрів на швидкість збіжності ігрового методу. Залежність порядку швидкості збіжності θ від параметрів α та β при роботі методу (5) у знакододатному середовищі зображена у вигляді стовпчикової діаграми на рис. 2.

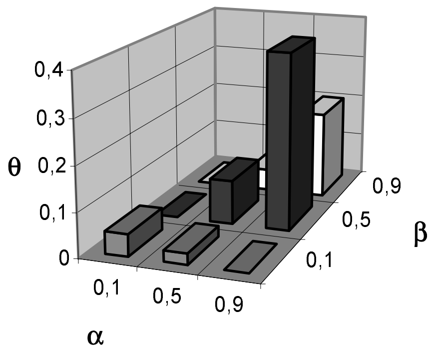

Рис. 2. *Порядок швидкості збіжності методу (5) у знакододатному середовищі для усередненої по реалізаціях похибки доповняльної нежорсткості.*

Значення параметра θ визначено як середнє по 100 реалізаціях випадкового процесу Δ_n.

З отриманих результатів видно, що для вибраних дискретних значень параметрів найбільший порядок швидкості збіжності методу (5) досягається при $\alpha = 0.9$, $\beta = 0.5$, що добре узгоджується з теоретичними оцінками.

Висновки. Проведені теоретичні дослідження та результати комп'ютерного моделювання дають можливість встановити обмеження на параметри градієнтного ігрового методу (5), які забезпечують виконання умови доповняльної нежорсткості (4) у знакопостійних середовищах з імовірністю 1 або у середньоквадратичному. Враховуючи специфічний полілінійний вигляд функцій середніх виграшів (2), метод (5) може бути нестійким для розв'язків гри у повністю змішаних стратегіях. Для забезпечення стійкості розв'язків гри у змішаних стратегіях доцільно виконати регуляризацію градієнтного методу [3]. Визначення умов збіжності регуляризованого методу у знакопостійних середовищах потребує додаткового дослідження.

1. Батищев Д.И. Методы оптимального проектирования. - М.: Радио и связь, 1984. - 248 с. 2. Вазан М. Стохастическая аппроксимация. - М.: Мир, 1972. - 295 с. 3. Назин А.В., Позняк А.С. Адаптивный выбор вариантов: Рекуррентные алгоритмы. - М.: Наука, 1986. - 288 с.