Ryszard Zieliński, Peter Neumann

Stochastische Verfahren zur Suche nach dem Minimum einer Funktion

Akademie-Verlag, Berlin 1983, w serii Mathematische Forschung - Band 16, str. 133, ISSN 0138-3019.

W monografii przedstawione zostały losowe metody szukania ekstremum funkcji. W ostatnich latach, z uwagi na szeroki dostęp do komputerów oraz łatwość stosowania metod losowych, gwałtownie wzrosło zainteresowanie nimi. Powstało wiele ciekawych metod heurystycznych, niestety nie mających dostatecznego uzasadnienia teoretycznego. Książka ta stanowi jedną z nielicznych prób wypełnienia luki pomiędzy bogatą praktyką a skromną teorią losowych metod optymalizacji. Jej wyjątkowa wartość polega na oryginalnym i głębokim potraktowaniu problematyki poszukiwania ekstremum globalnego, problematyki szczególnie trudnej i takiej, gdzie stosowanie metod losowych jest najbardziej uzasadnione. Oto krótki przegląd treści.

W krótkiej (dwustronicowej) przedmowie autorzy wyjaśniają, że „nasz wykład staramy się utrzymać na poziomie matematycznym inżyniera", a „książka spełni swe zadanie, jeżeli pomoże praktykom w lepszym zrozumieniu metod, którymi tak chętnie się posługują i jeżeli przyciągnie ona probabilistów i statystyków o numerycznych zainteresowaniach do tej bogatej tematyki".

Rozdział I - „Sformułowanie zadania. Algorytmy losowe" (33 strony; 26% całej objętości). Jest to rozdział o charakterze wstępnym, w którym autorzy formułują zadanie optymalizacji w skończenie wymiarowej przestrzeni euklidesowej E^k oraz podają szereg przykładów algorytmów losowych służących do jego rozwiązania. Dokonany wybór jest głównie wynikiem zainteresowań obu autorów, jednakże oddaje on dość wiernie obraz całości rozważanej problematyki. Algorytmy zostały podzielone na cztery kategorie: algorytmy z losowaniem kierunku, algorytmy z estymacją gradientu, algorytmy z losowaniem nowych punktów oraz algorytmy adaptacyjne.

Autorzy dokonują klasyfikacji algorytmów ze względu na nośnik generowanych punktów losowych w kolejnych krokach iteracyjnych. I tak algorytmy wymagające generowania punktów losowych w „bliskim sąsiedztwie" najlepszego z dotychczasowych wyników noszą nazwę algorytmów lokalnych, natomiast w przypadku, gdy nośnikiem generowanych punktów losowych jest cały zbiór rozwiązań dopuszczalnych mówimy o algorytmach globalnych.

W dodatku do rozdziału I podano sposoby generowania punktów losowych o rozkładzie jednostajnym na sferze w skończenie wymiarowej przestrzeni euklidesowej E^k.

Rozdział II - „Algorytmy lokalne" (24; 19%) - jest poświęcony badaniu efektywności algorytmów lokalnych. Zachowanie się tych algorytmów zanalizowano w oparciu o wprowadzone miary lokalnej efektywności w „polu liniowym" i w „polu kwadratowym". Miary te dają pewne porównanie danego algorytmu losowego z najprostszymi, deterministycznymi algorytmami gradientowymi w przypadku minimalizacji (odpowiednio) funkcji liniowej lub kwadratowej. Przykładowo, efektywność w polu liniowym zdefiniowano jako

$$(1) \qquad l_k(A) = \frac{k+1}{m_n} E \frac{(f_X(X^n), \xi^n)}{\| f_X(X^n) \|},$$

gdzie $X^n \in R^k$ jest n-tym przybliżeniem minimum, f_X - gradientem funkcji minimalizowanej, ξ^n jest wersorem kierunku $X^{n+1} - X^n$ wygenerowanego przez algorytm losowy, $(\cdot,\cdot)$ oznacza iloczyn skalarny, m_n jest liczbą obserwacji niezbędnych do wygenerowania kierunku ξ^n (po znalezieniu się w punkcie X^n), a k+1 jest najmniejszą liczbą obserwacji pozwalającą wyznaczyć wartość gradientu funkcji liniowej.

Ponadto w rozdziale tym na przykładowo wybranym algorytmie zilustrowano zagadnienie zbieżności z prawdopodobieństwem 1 losowych algorytmów lokalnych (tw. 7 str. 60).

Rozdział III - „Algorytmy globalne" (34; 26%). W rozdziale tym autorzy przedstawili prostą metodę Monte Carlo (np. w literaturze anglosaskiej znaną jako pure random search) oraz omówili zagadnienie zbieżności z prawdopodobieństwem 1 wybranych algorytmów globalnych, a mianowicie prostej metody Monte Carlo i jej pewnych uogólnień. Ponadto podali przegląd metod heurystycznych szukania ekstremum globalnego oraz omówili następujące problemy pokrewne:

- estymacja mody gęstości rozkładu prawdopodobieństwa,
- estymacja najmniejszej wartości funkcji,
- podejście bayesowskie.

Rozdział IV - „Zagadnienie wieloekstremalne" (6; 5%). W rozdziale tym omówiono zagadnienie jednoczesnego wykrycia wszystkich minimów lokalnych funkcji wieloekstremalnej. Przedstawiono dwa sposoby podejścia do tego zadania. Pierwszy sposób polega na określeniu liczności próby niezbędnej do wykrycia z zadanym prawdopodobieństwem wszystkich minimów o „dużych obszarach przyciągania". Drugi sposób polega na sprowadzeniu zadania do wyznaczenia optymalnej, bayesowskiej reguły zatrzymania procesu szukania przy funkcji straty będącej średnią ważoną kosztów obserwacji i kosztów związanych z niewykryciem wszystkich minimów lokalnych.

Rozdział V - „Losowe metody poszukiwania minimum funkcji regresji" (14; 11%). Rozdział ten zawiera zwięzłe omówienie własności oraz przykładowe twierdzenia o zbieżności podstawowych metod poszukiwania minimum globalnego funkcji regresji.

Przedstawione metody są oryginalnymi modyfikacjami prostej metody Monte Carlo i jej uogólnień, zwiększającymi efektywność w przypadku, gdy minimalizowana funkcja obserwowana jest z błędem losowym.

Przedstawimy teraz uwagi szczegółowe oraz komentarze do kolejnych rozdziałów recenzowanej monografii.

Rozdział I ma charakter wprowadzający i jako taki nie wymaga szerszego komentarza. Zaproponowany podział algorytmów losowych, jak każdy tego typu podział jest w pewnej mierze arbitralny. Warto może zwrócić uwagę na pewne jego niekonsekwencje. Na przykład idea najszybszego spadku (algorytm 1E) może być i jest stosowana także w przypadku algorytmów gradientowych - trudno zatem zrozumieć jej omówienie w podrozdziale dotyczącym algorytmów z losowaniem kierunku.

Ponadto niezrozumiałe jest określenie istotnego kresu dolnego minimalizowanej funkcji względem dowolnego rozkładu na jej dziedzinie, a nie np. względem miary Lebesgue'a.

Zasadniczym celem rozdziału II jest udzielenie odpowiedzi na pytanie „czy i ewentualnie, ile tracimy zastępując dokładne metody deterministyczne losowaniem kolejnego przybliżenia »na chybił trafił«". Autorzy ograniczają się do rozważenia tego pytania w przypadku minimalizacji funkcji liniowej (co mniej więcej odpowiada badaniu zachowania się algorytmów „z dala od minimum lokalnego") oraz funkcji kwadratowej $\|x\|^2$ (co w zasadzie odpowiada analogicznej analizie funkcji lokalnie symetrycznej „w pobliżu jej punktu stacjonarnego"). Autorzy formułują odpowiedzi zasadniczo słuszne, ale argumentacja do nich prowadząca nie jest w pełni zadowalająca. Przykładowo, twierdzenie 6(iii) orzeka, że w polu kwadratowym algorytm losowy 2A ma tę samą efektywność co deterministyczny algorytm gradientowy, trudno więc wyciągnąć stąd wniosek, że „algorytmy losowe trudno jest oceniać jako algorytmy mogące skutecznie konkurować z metodami deterministycznymi" (por. str. 57, ostatni akapit punktu 2.2.1). Generalnie zaś przyczyną takiego stanu rzeczy jest oparcie się na jednym tylko i to mocno kontrowersyjnym typie miary efektywności algorytmów (por. podany wyżej wzór (1)). Zaproponowana miara lokalnej efektywności, stosowana zresztą od początku lat 60-tych, jest kontrowersyjna z trzech powodów. Po pierwsze, metodologicznie słuszniejsze byłoby badanie rzutu kierunku wylosowanego ξ^n na kierunek $X^n - X^*$ (gdzie X^* oznacza położenie minimum), a nie na kierunek gradientu $f_X(X^n)$ (nie ma to znaczenia tylko dla funkcji liniowych oraz funkcji $a\|X\|^2$, $a > 0$). Po drugie, sposób uzależnienia efektywności od liczby niezbędnych obserwacji - odpowiednio m_n i $k+1$ w przypadku miary (1) - jest oczywiście całkowicie arbitralny. Po trzecie wreszcie, zgodziwszy się, że warto badać dokładność „odtwarzania" przez wersor ξ^n kierunku gradientu $f_X(X^n)$, trudno zaakceptować sposób oceny tej dokładności. Autorów interesuje jedynie wartość oczekiwana kosinusa kąta między ξ^n a $f_X(X^n)$, $\cos(\xi^n, f_X(X^n))$. Ze statystycznego punktu widzenia należałoby skonstruować jakąś funkcję straty związaną z błędem „odtwarzania" kierunku gradientu, np. rozważyć błąd średniokwadratowy $E[\cos(\xi^n, f_X(X^n)) - 1]^2$. W ten sposób uwzględniono by także wariancję błędu nota bene recenzenci nie zgadzają się ze stosowaną dość często ideą uwzględniania tejże wariancji, polegającą na dzieleniu przez nią efektywności (1) - por. np. Rubinstein (1981), rozdz. 7). Trzeba tu jeszcze zaznaczyć, że arbitralność uwzględniania liczby stosowanych obserwacji jest nie do uniknięcia przy zajmowaniu się tylko lokalnymi ocenami efektywności. Unika się zaś jej przechodząc do ocen globalnych, najbardziej zresztą interesujących użytkownika, a w omawianej monografii pominiętych. Oceny takie (sformułowane np. dla odległości średniokwadratowej) odpowiadają na pytania następujące:

- startując z zadanego punktu X^1, ilu potrzeba obserwacji wartości funkcji, aby znaleźć się w zadanej odległości od punktu X^*,

- startując z zadanego punktu X^1, jak blisko znajdziemy się punktu X^* po dokonaniu zadanej liczby obserwacji.

Co więcej, w przeciwieństwie do ocen lokalnych, oceny globalne pozwalają sformułować problem optymalnego doboru parametrów algorytmu, w szczególności współczynnika długości kroku.

Przy lekturze rozdz. II recenzentom nasunęły się jeszcze następujące uwagi szczegółowe.

- Metody lokalne pozostają zbieżne w przypadku funkcji słabo różniczkowalnych (oczywiste) oraz półciągłych (wskazuje na to idea tzw. uogólnionych gradientów stochastycznych) - por. przedostatni akapit na str. 42.
- Tytuł paragrafu 2.1 nie jest dobry - metody omawiane w paragrafie 2.2 to również metody z ustaloną długością kroku.
- Komentarz do definicji 3 jest nieco mylący - nie definiuje się tam oczywiście obciążenia w sensie statystycznym.
- W definicji 4 kontrowersyjne jest umieszczenie w mianowniku wielkości $\| f_X(X^n) \|^2$, a nie $\| f_X(X^n)\| \cdot \| \hat{f}_X(X^n) \|$.
- Szacowanie wielkości

$$\sum_{i=0}^{m-1} v_i u_i^{2m+1}$$

(por. str. 58-59) jest niepotrzebne. Jej dokładna wartości wynosi

$$(-1)^{m-1} 2^{-1} \prod_{i=0}^{m-1} u_i^2$$

(por. Fabian (1968), wzór 3.3.1); nota bene autorzy nie zaznaczają, że metoda omówiona w punkcie 2.2.2, poza pomysłem losowania indeksu J należącym do Zielińskiego, pochodzi od Fabiana.

- Omówienie problemu zbieżności z prawdopodobieństwem 1 na przykładzie algorytmu (2.3.2) tylko jest o tyle niewystarczające, że algorytm ten ma charakter teoretyczny raczej niż praktyczny (jak w praktyce konstruować ciąg γ_n?!).
- Teza c) twierdzenia 7 (str. 61) jest błędnie sformułowana; najlepiej byłoby zapisać ją następująco:

$$\limsup_{n\to\infty} n\left[f(X^n) - f(X^*)\right] \leqslant \frac{2kL\eta^2}{\gamma} \qquad \text{z p.1.}$$

Autorzy wpadli tu chyba we własną pułapkę źle odczytując swoje, niezbyt fortunne sformułowanie. Autorzy piszą zawsze „z prawdopodobieństwem 1 istnieje taka liczba naturalna N, że ...", zamiast wyraźniej i już na pewno jednoznacznie „dla każdego ω ze zbioru o mierze 1 istnieje taka liczba N, że ...".

Skoncentrowanie się w rozdziale III na prostej metodzie Monte Carlo i jej uogólnieniach jest ze wszech miar uzasadnione. Metody te są zbieżne z prawdopodobieństwem 1 przy braku jakichkolwiek założeń o funkcji minimalizowanej (poza jej mierzalnością) i dają wgląd w całość problematyki szukania minimum globalnego. Przegląd metod heurystycznych i zagadnień pokrewnych jest bardzo dobry i ciekawy. Może tylko warto by jeszcze wspomnieć (ze względu na ich popularność) o metodach związanych z ideą tzw. funkcjonałów wygładzających oraz (ze względu na ich teoretyczny fundament) o metodach wykorzystujących znaną charakteryzację punktu X^*; co prawda wystarczyłoby tu chyba odesłanie czytelnika do literatury (np. Rubinstein (1981)).

Rozdział IV jest pierwszą w literaturze światowej i bardzo ciekawą (należącą do R. Zielińskiego) próbą porządnego i ogólnego sformułowania problemu szukania minimum globalnego, uwzględniającą to, że liczba wszystkich obserwacji funkcji musi być ograniczona.

Zamieszczenie w monografii rozdziału V, poświęconego szukaniu minimum globalnego funkcji regresji i, zarazem, pominięcie problemu poszukiwania minimum lokalnego funkcji regresji może się wydać kontrowersyjne. Wobec jednak obfitości prac dotyczących problemów lokalnych i braku w literaturze światowej porządnego wprowadzenia do problematyki optymalizacji globalnej, zamieszczenie w monografii wprowadzenia było nader pożądane. Czytelnika ambitniejszego może tylko nieco rozczarować duża zwięzłość wykładu, skądinąd systematycznego i zgodnego z duchem reszty monografii (bazującego w głównej mierze na oryginalnych wynikach R. Zielińskiego). Z praktycznego punktu widzenia odczuwa się brak bardziej szczegółowego opisu i analizy algorytmów adaptacyjnych (dobór rozkładów Q_n!).

Reasumując: zważywszy, że zastosowania algorytmów losowych do szukania minimum globalnego są znacznie ważniejsze od ich ewentualnego stosowania w problemach optymalizacji lokalnej, o wartości książki powinny decydować jej rozdziały III - V. Autorzy uczynili zatem trafnie, traktując rozdziały I i II jako wprowadzające, choć szkoda oczywiście, że analiza algorytmów lokalnych nie jest dość głęboka. Rozdziały III - V pozwalają zaś monografię jako całość ocenić bardzo wysoko.

Komentując całą książkę, wypada jeszcze zwrócić uwagę na zawsze bardzo dobry wybór przykładów ilustrujących istotę rozważanych problemów oraz na celowość umieszczenia na końcu każdego rozdziału uwag bibliograficznych, wskazujących, skąd zaczerpnięto zreferowane wyniki oraz jakie wyniki czy też ujęcia pominięto. Książka napisana jest starannie - usunięcie nielicznych błędów korektorskich nie sprawi czytelnikowi żadnych trudności. Krytycznie natomiast ocenić należy nadmierną szczegółowość dowodów.

Chociaż monografia ma stanowić wprowadzenie do teorii algorytmów losowych, wydaje się, że zamieszczenie kilku przykładów numerycznych (lub symulacyjnych tam, gdzie trudno o ścisłe porównanie różnych algorytmów) podniosłoby jeszcze atrakcyjność książki.

LITERATURA CYTOWANA

V. Fabian, On the choice of design in stochastic approximation methods, Ann. Math. Statist. 39 (1968), 457-465.

R.Y. Rubinstein, Simulation and the Monte Carlo method, Wiley, New York 1981.

JACEK KORONACKI

ANDRZEJ SIEROCIŃSKI