ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1976 год Выпуск 67

УДК 515.251.26+519.3:330.115

ОБ ОДНОМ АЛГОРИТМЕ ПОИСКА ГЛОБАЛЬНОГО ЭКСТРЕМУМА ФУНКЦИИ

Г.С.Лбов, А.А.Трунов

Предлагается алгоритм приближенного отыскания глобального экстремума функции, удовлетворяющей условию Липшица. Оценка константы Липшица определяется в ходе поиска. Эффективность работы алгоритма проверена на ряде модельных примеров.

Пусть некоторая функция $F(x)$ определена на параллелепипеде D n-мерного евклидова пространства. Известно, что функция $F(x)$ удовлетворяет на D условию Липшица и в произвольной точке $x \in D$ может быть определено значение функции путем либо эксперимента, либо вычисления. Такую пару $\langle x, F(x) \rangle$ будем называть экспериментом или шагом поиска.

В ходе поиска разрешено провести фиксированное число T экспериментов (считается, что определение значения функции связано с большими затратами на эксперимент, поэтому число T обычно невелико). Правило перестановки T точек $x^1, \ldots, x^i, \ldots, x^T$ ($x^i \in D$, $i = 1, \ldots, T$), в которых определяется значение функции $F(x)$, назовем стратегией поиска.

Пусть, для определенности, отыскивается глобальный максимум функции. Ясно, что за фиксированное число шагов максимум функции может быть не найден. Поэтому за приближенное значение максимума принимается наибольшее значение функции, достигнутое за T шагов.

Таким образом, в ходе поиска минимизируется величина $\Delta = F^* - F_{max}$, где F^* – значение глобального максимума функции, F_{max} – наибольшее значение функции, достигнутое за T шагов.

§I. Описание алгоритма

Предположим, что область D есть единичный гиперкуб n-мерного евклидова пространства ($0\le x_j\le 1$ для $j=1,\dots,n$). Если D – гиперпараллелепипед, то линейным преобразованием переходим к гиперкубу.

При построении алгоритма используется условие Липшица:

$$|F(x) - F(y)|\le L|x-y|;\quad x,y\in D.$$

Первая точка ставится случайно в области D. Дальнейшие эксперименты проводятся по следующей схеме. Пусть проведено t экспериментов ($0< t< T$), т.е. определены пары $\langle x^i, F(x^i)\rangle$, $i=1,\dots,t$. Необходимо указать точку ($t+1$) -го эксперимента – x^{t+1}. Для этого на области D определяется функция $G(x)$, задающая верхнюю границу для значений функции $F(x)$,

$$G(x) = \min_{i=1,\dots,t} q(x^i, x),$$

где $q(x^i,x)=F(x^i)+L|x^i-x|$, $x\in D$. Эксперимент ($t+1$) проводится в точке максимума функции $G(x)$. Такая стратегия адаптированного планирования экспериментов предложена в работе [I].

Однако при использовании указанной стратегии для многомерного пространства возникают трудности в отыскании максимума функции $G(x)$. Для определения максимума $G(x)$ был использован метод Монте-Карло, дающий для рассматриваемого случая вполне приемлемое решение. Необходимо небольшое число испытаний (вычислений функции $G(x)$ в случайно выбранных точках), чтобы получить удовлетворительное приближенное решение. Например, за IOO испытаний с вероятностью 0,95 попадаем хотя бы один раз в трехпроцентную окрестность максимума ункции $G(x)$. Под M-окрестностью максимума некоторой функции $f(x)$, $x\in D$, понимается область $D(M)=\{x: f(x)>M\}$, где $\min f(x)\le M\le \max f(x)$. Под R-процентной окрестностью понимается область $D(M)$ такая, что $\frac{V_{D(M)}}{V_D}\cdot 100\% = R$, где $V_{D(M)}$ и V_D – объемы областей $D(M)$ и D. Легко видеть, что для функции $G(x)$ объем $V_{D(M)}$ меняется непрерывно. Поэтому R-процентная окрестность всегда существует для произвольных R.

Кроме того, при решении модельных задач было замечено, что для повышения эффективности работы алгоритма в случае многомерного пространства и небольшого числа T желательно учитывать так называемый "эффект границы", приводящий к необходимости введения на границе области D поправки Δq к функции $q(x^i,x)$. В противном случае после проведения первого эксперимента последующие будут проводиться только на границе (либо близко к границе) области D.

Введение поправки Δq делается из следующих эвристических соображений. Обозначим через F^t_{max} максимальное значение функции, полученное за t шагов. Тогда очевидно, что в области D_i, представляющей n-мерный шар с радиусом

$$\rho_i=\frac{F^t_{max} - F(x^i)}{L}$$

и с центром в точке x^i, проводить эксперименты не следует, так как всегда $F(x)\le F^t_{max}$ для $x\in D_i$ [2]. Назовем область D_i областью "влияния" i-го эксперимента $\langle x^i, F(x^i)\rangle$. Если значения функции для двух разных экспериментов $\langle x^i, F(x^i)\rangle$ и $\langle x^l, F(x^l)\rangle$ равны, то и объемы областей "влияния" должны быть равны ($V_{D_i} = V_{D_l}$). Однако если одна из этих точек (например, x^l) оказывается на границе (либо близко к границе) гиперкуба, то только некоторая часть области "влияния" D_l оказывается внутри гиперкуба. Обозначим объем этой части области через $V_{D'}$. Область "влияния" такого же объема мы приписали бы эксперименту со значением функции $F(x')>F(x^l)$. В качестве поправки к функции $q(x^l,x)$ возьмем $\Delta q = F(x')-F(x^l)$. Исходя из построения областей "влияния", ясно, что на границах этих областей значения функции $q(x^l,x)$ равны. Отсюда

$$F(x^l) + L\rho_l = F(x') + L\rho(x'),$$

где $\rho(x')$ – радиус n-мерной сферы объема $V_{D'}$ и

$$\Delta q = F(x') - F(x^l) = L(\rho_l - \rho(x')).$$

Таким образом, учитывая поправку на границу, функция $q(x^l,x)$ имеет вид:

$$q(x^i,x) = F(x^i) + L|x^i - x| - L(\rho_i - \rho(x')).$$

Ясно, что при удалении точки эксперимента от границы внутрь области D $V_{D'} \to V_{D_i}$ и соответственно $\rho \to \rho_i$ и $\Delta q \to 0$.

Для вычисления ρ необходимо знать объем $V_{D'}$. Для упрощения вычислений вместо гипершара, задающего область D_i, возьмем гиперкуб с ребрами длины l, параллельными осям координат, и положим объем этого гиперкуба равным

$$V_{D_i} = \frac{V_D}{T} = \frac{1}{T}.$$

В этом случае радиус описанного гипершара

$$\rho_i = \frac{\sqrt{n}}{2} \sqrt[n]{V_{D_i}} = \frac{\sqrt{n}}{2\sqrt[n]{T}}.$$

Аналогично

$$\rho = \frac{\sqrt{n}}{2} \sqrt[n]{V_{D'}},$$

где

$$V_{D'} = \prod_{j=1}^{n} \left(\frac{1}{2} + \delta_j\right).$$

Величина δ_j определяется так:

$$\delta_j = 1 - x_j, \text{ если } x_j + \frac{1}{2} > 1;$$

$$\delta_j = x_j, \text{ если } x_j - \frac{1}{2} < 0;$$

$$\delta_j = \frac{1}{2} \text{ – в остальных случаях}.$$

§2. Выбор константы Липшица

Для того, чтобы решить прикладную задачу, необходимо знать константу Липшица L, которая, как правило, неизвестна. Возникает необходимость оценивать ее в ходе поиска. Пусть проведено t экспериментов. Для каждой пары экспериментов, например $\langle x^k, F_k\rangle$ и $\langle x^l, F_l\rangle$, оценивается константа Липшица

$$L_{kl} = \frac{|F_k - F_l|}{|x^k - x^l|}.$$

Из всех таким образом полученных констант Липшица определяется максимальная константа. Но эту константу можно считать лишь сильно заниженной оценкой истинной константы Липшица. Под истинной константой Липшица понимается величина

$$L = \sup_{x,y \in D} \frac{|F(x) - F(y)|}{|x - y|}.$$

Поэтому t-й эксперимент проводится с константой Липшица L_{t-1}, полученной следующим образом. На каждом шаге считается, что если в среднем за каждый эксперимент прирост константы Липшица составлял некоторое ΔL, то для оставшихся $T-t+1$ экспериментов будет точно такой же прирост на каждый эксперимент. Таким образом, можно оценить константу Липшица, которая получается через T шагов. Она и берется в качестве оценки константы при проведении t-го эксперимента. После постановки t-го эксперимента эта константа пересчитывается с учетом вновь пришедшей информации. Конкретно в алгоритме эта идея реализуется следующим образом:

1) первая точка выбирается случайно;

2) вторая точка выбирается с произвольной константой Липшица (например, $L = 10^3$), вычисляется L_{12} между первой и второй точками

$$L_{12} = L_2 = \frac{|F_2 - F_1|}{|x^2 - x^1|};$$

3) третья точка выбирается с произвольной константой Липшица ($L = 10^3$), вычисляется $L_3 = \max\{L_2, L_{13}, L_{23}\}$;

4) четвертая точка выбирается с константой Липшица

$$\bar{L}_4 = \frac{T}{3} L_3 - \frac{T-3}{3} L_2,$$

вычисляется $L_4 = \max\{L_3; L_{14}, L_{24}, L_{34}\}$ и т.д.;

5) точка t выбирается с константой Липшица

$$L_t = \frac{T}{t-1} L_{t-1} - \frac{T - t + 1}{t-1} L_2 ,$$

вычисляется $L_t = \max \{ L_{t-1};\ L_{1t}, L_{2t}, \dots, L_{t-1,t} \}$.

Необходимо отметить, что использование в алгоритме вместо истинного значения константы Липшица прогнозируемого значения, может привести к тому, что глобальный экстремум может попасть в область, запрещенную для дальнейшего проведения экспериментов.

§3. Проверка работоспособности алгоритма на модельных примерах

Пример I. Многоэкстремальная задача. Найти

$$I = \max_{0 \le x_i \le 1} e^{-4\rho} \cos 2\pi\rho, \text{ где } \rho = \sqrt{\sum_{i=1}^{2} (x_i - 0{,}5)^2}.$$

Решение задачи очевидно: $I = 1$. Было взято число экспериментов $T = 100$. Экстремум с точностью до трех знаков (0,999...) был найден на 28-м шаге.

Пример 2. Многоэкстремальная задача [4]. Найти

$$I = \min \{ (x^1)^2 + (x^2)^2 - \cos(18x^1) - \cos(18x^2) \}$$

на прямоугольнике

$$-0{,}25 \le x^1 \le 0{,}5;$$
$$-0{,}125 \le x^2 \le 0{,}625 .$$

Решение задачи: $I = -2$. Было взято число экспериментов $T = 100$. Хорошее приближение было получено на 17-м шаге ($I = -1{,}974$). Для сравнения в таблице (стр.75) приведены результаты решения этой же задачи другими методами.

Пример 3. Найти минимум той же самой функции на прямоугольнике

$$0{,}685 \le x^1 \le 17{,}975;$$
$$0{,}032 \le x^2 \le 17{,}162 .$$

Было взято число экспериментов $T = 100$. Хорошее приближение было достигнуто на 43-м шаге (-1,903).

Таблица

Число шагов	Методы			
	Метод Монте-Карло	О Б А*)	Метод 2*)	Предлагаемый алгоритм
10	-1,33	-1,33	-1,33	
30	-1,66	-1,68	-1,85	-1,974
40	-1,7	-1,85		-1,974
100	-1,75	-1,97		-1,974
150	-1,85			

Пример 4. Одноэкстремальная задача. Найти максимум функции правдоподобия

$$F(\mu, \sigma) = \prod_{i=1}^{6} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2},$$

где $\{x_1 = -1{,}39,\ x_2 = -0{,}68,\ x_3 = -0{,}21,\ x_4 = 0{,}21,\ x_5 = 0{,}68,\ x_6 = 1{,}39\}$ представляет собой выборку объема $N = 6$ из нормального распределения ($\mu = 0$ и $\sigma = 1$).

Поиск максимума функции $F(\mu,\sigma)$ осуществлялся на прямоугольнике

$$-1{,}39 \le \mu \le 1{,}39;$$
$$0{,}01 \le \sigma \le 2{,}78 .$$

Было взято $T = 50$. На 13-м шаге найдено приближенное значение максимума функции $F(\mu,\sigma)$, равное 0,00035 в точке $\mu = 0{,}013$ и $\sigma = 0{,}993$. (Аналитическое решение дает значение максимума, равное 0,00037 в точке $\mu = 0$, $\sigma' = \sqrt{0{,}8128} \approx 0{,}9$.)

Литература

I. ПИЯВСКИЙ С.А. Один алгоритм отыскания абсолютного экстремума. -"Журнал вычислительной математики и математической физики", 1972, т.12, № 4, с.888-896.

---

*) ОБА - метод, приведенный в работе [4]. Метод 2 - многомерный байесовский метод, основанный на использовании условных математических ожиданий и дисперсий [3].

2. ЕВТУШЕНКО Ю.Г. Методы поиска глобального экстремума. -В кн.: "Исследование операций". Вып.4. М., Изд-во ВЦ АН СССР, 1974.

3. МОЦКУС И.Б. О байесовских методах поиска экстремума. -"Автоматика и вычислительная техника", 1972, №3, с.10-21.

4. ЖИЛИНСКАС А.Г. Об одном упрощении одношагового байесовского метода. -В кн.: "Автоматизированное оптимальное проектирование инженерных объектов и технологических процессов".Горький, 1974, с. 71-78.

Поступила в ред.-изд.отд.
2 февраля 1976 года