

УДК 519.248

Оптимизация нескольких критериев с помощью случайного поиска. Туртия С.Б./Научное приборостроение. Электронно-ионная оптика. Л.:Наука, 1989, с. 116-121.

Даны основные понятия векторной оптимизации и краткое описание метода ветвящегося адаптивного случайного поиска (ВАСП), обоснование сходимости метода, представлен оптимальный выбор параметров этого метода. Лит. - 4 назв., ил. - 1.

ОПТИМИЗАЦИЯ НЕСКОЛЬКИХ КРИТЕРИЕВ
С ПОМОЩЬЮ СЛУЧАЙНОГО ПОИСКА

О с н о в н ы е п о н я т и я в е к т о р н о й о п т и м и з а ц и и
и к р а т к о е о п и с а н и е м е т о д а в е т в я щ е г о с я
а д а п т и в н о г о с л у ч а й н о г о п о и с к а (В А С П)

Пусть на множестве $X \subset R^n$ задан набор функций-критериев $f_1, f_2, \dots, f_s$, каждую из которых желательно минимизировать:

$$f_i(x) \rightarrow \min, \quad i = \overline{1, s}.$$

Задача векторной оптимизации определяется как задача поиска некоторого множества $E \subset X$ и его образа $\Pi = F(E) \subset R^s$, обладающих свойством

$$F(x^*) = \min_{x \in X} F(x), \quad \forall x^* \in E, \quad (1)$$

где $F(x) = \langle f_1(x), f_2(x), \dots, f_s(x) \rangle : X \rightarrow Y = Y_1 \times Y_2 \times \dots \times Y_s; f_i : X \rightarrow Y_i, i = \overline{1, s}$. При этом $\min$ в формуле (1) понимается в том смысле, что не существует $x' \in X$ тако-го, что вектор $F(x')$ предпочтительнее вектора $F(x^*)$, $\forall x^* \in E$.

Напомним понятия предпочтительности и несравнимости [1, 2]: векторная оценка $F(x^k)$ предпочтительнее оценки $F(x^l) : F(x^k) < F(x^l)$, если $\forall i = \overline{1, s} : f_i(x^k) \leq f_i(x^l)$, $\exists j : f_j(x^k) < f_j(x^l)$.

Если между двумя нетождественными оценками нельзя установить отношение предпочтения, то они считаются несравнимыми: $F(x^k) \bar{<} F(x^l)$. Множество E называется множеством эффективных решений, а множество $\Pi = F(E)$ - множеством Парето.

Известно [3], что в случае однокритериальной оптимизации направление, выбираемое по методу наилучшей пробы, при увеличении числа проб все более приближается к наилучшему из всех возможных направлений оптимизации. Поэтому на первом

этапе поиска методом ВАСП во всем X равномерно и независимо разыгрываются случайные точки $x^1, x^2, \dots, x^{M_k}$ и вычисляются соответствующие значения вектор-функции $F(x)$. Точка, дающая наиболее предпочтительную оценку или первую из несравнимых, считается первым приближением к E .

На втором и всех последующих этапах в X в целях адаптации выделяется перспективная область — интервал интенсивного поиска (ИИП) с центром в лучшей точке предыдущего этапа.

На k -м этапе ($2 \leq k \leq K$) поиск ведется в двух областях: ИИП (k) с плотностью $H(k) = P(k)/2q_k$ и в $X \setminus \text{ИИП}(k)$ с плотностью $h(k) = (1 - P(k))/(1 - 2q_k)$, где $P(k)$ — вероятность попадания в ИИП (k),

$$P(k) = R \cdot |0,25 - q_k| - R \cdot 0,25 + 1,$$

где R — коэффициент осторожности [1], а q_k — полуширина ИИП (k). При переходе к следующему этапу ИИП (k) сжимается

$$q_{k+1} = q_k / (1 + A/n), \quad q_1 = 0,5,$$

где A — коэффициент сжатия; n — размерность множества X . Сжатие происходит до достижения заданной точности ε . Таким образом, общее число этапов может быть вычислено по формуле

$$K = \left\lceil \frac{\ln(0,5/\varepsilon)}{\ln(1 + A/n)} \right\rceil + 1. \quad (2)$$

Для получения достаточно представительной выборки из множества Π , а также в целях достаточно адекватного решения задачи оптимизации овражных и сильно осциллирующих функций применяется процедура ветвления процесса поиска [1]: с первого и до последнего шага поиск ведется параллельно и независимо по нескольким направлениям. Если в процессе поиска две ветви "слипаются", то одна из них обрывается и взамен ее порождается новая.

Сходимость метода

Пусть множество Y обладает свойством юго-западной замкнутости, тогда в соответствии с теоремой 1 в работе [4], $\Pi \neq \emptyset$, следовательно, $\forall y \in Y$ можем определить

$$\rho(y, \Pi) = \min_{y_n \in \Pi} \rho(y, y_n). \quad (3)$$

Пусть y^k, y^{k+1} — точки, построенные какой-либо ветвью алгоритма ВАСП, соответственно на k и $(k+1)$ -м этапах.

Утверждение 1. $\rho(y^{k+1}, \Pi) \leq \rho(y^k, \Pi)$, т.е. метод ВАСП является релаксационным, и на каждом этапе алгоритма происходит уточнение результата полученного на предыдущем этапе.

Доказательство. Рассмотрим процесс построения y^{k+1} . На $(k+1)$ -м этапе в X в соответствии с плотностями $H(k+1), h(k+1)$ разыгрывается M_k точек $x^1, x^2, \dots, x^{M_k}$ и вычисляются $y^i = F(x^i), i = \overline{1, M_k}$. Если среди них нет точек предпочтительнее чем y^k , то доказываемое утверждение очевидно, так как в этом случае $y^{k+1} = y^k$. Однако, если M_k достаточно велико, то с большой вероятностью лучшая точка из $\{y^i\}_{i=\overline{1, M_k}}$ (обозначим ее y^{k+1}) предпочтительнее точки y^k (это следует из вычислительной схемы алгоритма ВАСП.)

Для y^k в соответствии с (3) $\exists y_n^k \in \Pi: \rho(y^k, y_n^k) = \rho(y^k, \Pi)$. Но $y^{k+1} \prec y^k$, следовательно [1], $\exists i \in \{1, 2, \dots, n\}: y_i^{k+1} < y_i^k$. Из совокупности свойств расстояния и множества Парето следует, что $\rho(y^{k+1}, y_n^k) < \rho(y^k, y_n^k) = \rho(y^k, \Pi)$. Но

$$y_n^k \in \Pi \Rightarrow \rho(y^{k+1}, y_n^k) \geq \rho(y^{k+1}, \Pi).$$

Отсюда $\rho(y^{k+1}, \Pi) < \rho(y^k, \Pi)$. Утверждение доказано.

Рассмотрим поведение последовательности $\{x^k\}$, построенной алгоритмом ВАСП ($x^k = F^{-1}(y^k)$), в случае компактного множества X . Траектория одной ветви поиска может быть представлена как некоторый итеративный процесс, аналогичный одномерным релаксационным алгоритмам:

$$x^{k+1} = x^k + \eta(x^k, \xi^k, X) \xi^k, \quad (4)$$

где $\eta(a, b, X) = \begin{cases} 1, & a+b \in X \wedge F(a+b) < F(a) \\ 0, & a+b \notin X \vee F(a) < F(a+b) \vee F(a) \approx F(a+b), \end{cases}$
 ξ^k — случайный вектор распределен с плотностями $H(\xi)$ и $h(\xi)$.

Введем функцию $g(x)$, обладающую свойствами: $g(x) \geq 0, \forall x \in X; g(x) < g(y)$, если $F(x) < F(y); g(x) = g(y)$, если $F(x) \approx F(y)$.

Функция $g(x)$ полностью определена в X и, согласно формуле (4), последовательность $g(x^k)$ является неотрицательным супермартингалом относительно неубывающей последовательности σ -алгебр, порожденных случайными величинами $g(x^1), \dots, g(x^{k-1})$. Следовательно, почти наверное существует $\lim_{k \rightarrow \infty} g(x^k) = g^* \geq 0$. Таким образом, чтобы доказать, что последовательность $y^k = F(x^k)$ сходится к некоторой не-улучшаемой точке $y^* = F(x^*)$ в указанном смысле, достаточно доказать, что $g^* = 0$ почти наверное.

Рассмотрим пространство случайных событий $\Omega = \{\omega\}$, определяющих поведение последовательности $\{x^k\}$ в множестве X . Запись $x^k(\omega)$ будет обозначать, что совокупность случайных событий ω вызвала переход из точки x^{k-1} в точку x^k . Траектории $x^k(\omega)$ в X , начиная с какого-то шага k_0 могут целиком находиться в некотором множестве $D \subset X$. В этом случае обозначим

$$\Omega_D(k_0) = \{\omega \in \Omega: x^k(\omega) \in D, \forall k \geq k_0\}. \quad (5)$$

В работе [4] для обоснования сходимости одномерных релаксационных алгоритмов введено понятие ε -возвратного множества. В многомерном случае это понятие будет аналогичным: множество $D, \subset X$ называется ε -возвратным для $D \subset X$, если траектории из (5) за конечное время пересекаются с D . Другими словами, если $D_1 \cap D \neq \emptyset$ и $\forall k_0 > 0, \exists \varepsilon > 0$:

$$\lim_{k \rightarrow \infty} P\{\{\omega: x^k(\omega) \in D_1\} \cap \Omega_D(k_0)\} \geq \varepsilon \cdot P\{\Omega_D(k_0)\},$$

где вероятность события A — $P\{A\}$.

Введем функцию на множестве Ω : $\eta(x^k, \beta) = P\{\omega: g(x^k) - g(x^{k-1}) \leq \beta\}$, соответствующую вероятности того, что уменьшение функции $g(x)$ на k -й итерации поиска будет не меньше, чем $-\beta$.

Функции $g(x)$, $\eta(x, \beta)$ и понятие ε -возвратности аналогичны введенным в [3] для доказательства теоремы 4.2.1. о сходимости одномерных релаксационных алгоритмов. Многомерный аналог данной теоремы будет выглядеть следующим образом.

Утверждение 2. Пусть $x^0 \in D \subset X; \exists D, \subset X$, являющееся ε -возвратным для D ; $\exists \beta < 0: \inf_{x \in D} \eta(x, \beta) > 0$, тогда последовательность $\{x^k\}$, порожденная соотношениями (4) почти наверное за конечное время выходит из множества D .

Доказательство данного утверждения почти целиком дублирует доказательство теоремы 4.2.1 из [3], поэтому приводить его не будем. Напомним, что все предыдущие рассуждения проводились в условиях компактности множества X . Пусть, кроме того, множество Y замкнуто на юго-западе (понятие юго-западной замкнутости — КЗЗ введено в работе [4]), а на вектор-функцию $F = \langle f_1, \dots, f_s \rangle$ по-прежнему никаких условий не наложено. В соответствии с (3) для $\forall x^0 \in X \exists \rho_0 = \rho(F(x^0), \Pi)$,

$$y^0 \in \Pi: \rho(F(x^0), y^0) = \rho_0.$$

Пусть x^0 — начальная точка поиска по методу ВАСП и $0 < \varepsilon \leq \rho_0$. Определим следующие множества:

$$\begin{aligned} D(\varepsilon) &= \{x \in X: \varepsilon \leq \rho(F(x), \Pi) \leq \rho_0\} \\ G(\varepsilon) &= \{x \in X: \rho(F(x), \Pi) < \varepsilon\}. \end{aligned} \quad (6)$$

Если $x^0 \notin E$, то $D(\varepsilon) \neq \emptyset$ (это следует из утверждения 1), а так как Y по условию обладает свойством КЗЗ, то по теореме 1 из [4] $\Pi \neq \emptyset$, вследствие этого, $G(\varepsilon) \neq \emptyset$. В соответствии с утверждением 1 для любых членов n последовательности $\{x^k\}$, порожденных алгоритмом ВАСП по формуле (4), справедливо $\rho(F(x^n), \Pi) \leq \rho(F(x^{n-1}), \Pi) \leq \dots \leq \rho(F(x^0), \Pi) = \rho_0$, поэтому $\{x^k\} \subset D(\varepsilon) \cup G(\varepsilon)$.

Утверждение 3. Пусть $x^0 \in D(\varepsilon)$; $\forall \varepsilon > 0 \exists D_\varepsilon \subset X$, являющееся возвратным для $D(\varepsilon)$; $\forall \varepsilon \in (0, \rho_0) \exists \beta < 0: \inf_{x \in D_\varepsilon} \eta(x, \beta) > 0$

Тогда последовательность $\{F(x^k)\}$, порожденная методом ВАСП, почти наверное сходится к множеству Парето Π .

Доказательство. Поскольку мы находимся в условиях утверждения 1, то последовательность $\{x^k\}$ почти наверное за конечное время выходит из множества $D(\varepsilon)$, а так как $\{x^k\} \in D(\varepsilon) \cup G(\varepsilon)$, то она окажется в множестве $G(\varepsilon)$, и, вследствие произвольного выбора $\varepsilon > 0$, это означает суть утверждения 3.

Таким образом, сходимость метода доказана.

Следующая теорема дает некоторую оценку скорости сходимости метода ВАСП.

Утверждение 4. Пусть множество Парето $\Pi \neq \emptyset$ и $\varepsilon > 0$, тогда

$$P\{\{x_i^k\}_{i=\overline{1, M_k}} \cap G(\varepsilon) \neq \emptyset / \omega_k\} \geq 1 - (\varepsilon_k)^{\overline{m}_k}, \quad (7)$$

где $1 \leq k \leq K$ из (2); $\{x_i^k\}_{i=\overline{1, M_k}}$ — множество исследуемых на k -м этапе точек, событие ω_k состоит в том, что m_k из них попадут в ИИП (κ); $\overline{m}_k = M_k - m_k$; множество $G(\varepsilon)$ определено по формуле (6), а точное значение зависящей от k бесконечно малой величины ε_k будет определено в ходе доказательства.

Доказательство. Обозначим выражение в левой части утверждения (7) через P и пусть событие ω_k имеет место. Из формул для $H(k)$ и $h(k)$ можно вывести, что $m_k = [(R \cdot |0,25 - q_k| - R \cdot 0,25 + 1) \cdot M_k]$.

Рассмотрим два случая:

1. ИИП (κ) $\cap G(\varepsilon) = G_k(\varepsilon) \neq \emptyset$. Объем множества $G_k(\varepsilon)$ обозначим через V_k , а множества $G(\varepsilon)$ через V_ε и рассмотрим совокупность событий

$$\begin{aligned} B_1 &= \{\omega: \{x_i^k(\omega)\}_{i=\overline{1, m_k}} \in G_k(\varepsilon)\}, \\ B_2 &= \{\omega: \{x_i^k(\omega)\}_{i=\overline{m_k+1, M_k}} \in G(\varepsilon) \setminus G_k(\varepsilon)\}. \end{aligned}$$

С одной стороны, $B_1 \cup B_2 = \{\omega: \{x_i^k(\omega)\}_{i=\overline{1, M_k}} \in G(\varepsilon)\}$, с другой — события B_1 и B_2 с очевидностью независимы. Поэтому

$$P = P_1 + P_2 - P_1 \cdot P_2,$$

$$\text{где } P_1 = P\{B_1\} = 1 - (1 - V_k / (2q_k)^n)^{\overline{m}_k}; \quad P_2 = P\{B_2\} = 1 - \left(1 - \frac{V_\varepsilon - V_k}{1 - (2q_k)^n}\right)^{\overline{m}_k}.$$

Так как $0 \leq \frac{V_\varepsilon - V_k}{1 - (2q_k)^n} \leq 1$, положим $\varepsilon_k = 1 - (V_\varepsilon - V_k) / (1 - (2q_k)^n)$. Подставив значения P_1 и P_2 в выражение для P и приняв во внимание, что $0 < \frac{V_k}{(2q_k)^n} \leq 1$, получим $P \geq 1 - (\varepsilon_k)^{\overline{m}_k}$.

2. ИИП (κ) $\cap G(\varepsilon) = \emptyset$, тогда $P = 1 - (1 - V_\varepsilon / (1 - (2q_k)^n))^{\overline{m}_k} \geq 1 - (\varepsilon_k)^{\overline{m}_k}$, так как $V_\varepsilon \geq V_k \geq 0$. Утверждение доказано.

Рассмотрим одну ветвь поиска. Из вычислительной схемы метода ясно, что свободных параметров два: A — коэффициент сжатия ИИП и M_k — число точек, исследуемых на k -м этапе (рассмотрим случай $M_1=M_2=\dots=M$). Экспериментальным путем были получены такие оптимальные границы выбора: $0,5 \leq A \leq n$; $10 \leq M \leq 100$ (решались тестовые задачи, в которых $2 \leq n \leq 10$).

Следует заметить, что если параметр A определяет весь ход поиска, его стратегию, то параметр M определяет тактику поиска на каждом конкретном этапе. От величины A в большей степени зависит, найдет ли алгоритм быстрый и устойчивый путь к множеству Парето (в одномерном случае — достигнет ли алгоритм зоны притяжения глобального минимума, или же он сойдется в локальном минимуме, имеющем большую зону притяжения). От величины M существенно зависят значения вероятностей притяжения локальных минимумов на данном конкретном этапе. Увеличение значения M , как правило, увеличивает точность приближения к множеству Парето, однако ведет к существенному возрастанию затрат машинного времени.

Рассмотрим множество центров интервалов интенсивного поиска $\{x^k\}_{k=\overline{1,K}}$, $x^k = F^{-1}(y^k)$; y^k — точка, признанная лучшей из построенных на k -м этапе, а число этапов K вычислено по формуле (2). Пусть множество эффективных решений $E \neq \emptyset$. Тогда аналогично формуле (3) можем ввести $\rho_k = \rho(x^k, E) = \min_{x \in E} \rho(x^k, x)$, и задача скорейшего выхода на множество Парето может быть сведена к задаче

$$\sum_{k=1}^K \rho_k \rightarrow \min. \quad (8)$$

Пусть m_k и $\bar{m}_k$ — те же, что были введены при доказательстве утверждения 4, а именно: m_k ($\bar{m}_k$) — среднее значение числа точек, попавших в ИИП (k) (в дополнение к ИИП (k)). Введем также случайные величины, характеризующие расстояние между двумя соседними точками соответственно в ИИП (k) и в $X \setminus \text{ИИП} (k)$. Легко можно получить оценки средних значений этих случайных величин:

$$\ell_k = \frac{2g_k}{\sqrt{m_k}} \cdot \sqrt{n}; \quad \bar{\ell}_k = \sqrt{\frac{1-(2g_k)^2}{\bar{m}_k}} \cdot \sqrt{n}.$$

Рассмотрим соотношение между ρ_1 и ρ_2 . Вследствие утверждения 1 $\rho_2 \leq \rho_1$. С другой стороны, $\rho_1 \leq \rho_2 + \rho(x^1, x^2)$.

Применив аналогичные рассуждения ко всем $\{x^k\}_{k=\overline{1,K}}$, получим $\sum_{k=1}^K \rho_k \leq \rho_K + \sum_{k=1}^{K-1} k \cdot \rho(x^k, x^{k+1})$.

Поскольку цель оптимизации — выход на множество Парето, величиной ρ_K можно пренебречь. Тогда задача (8) переходит в задачу

$$\sum_{k=1}^{K-1} k \cdot \rho(x^k, x^{k+1}) \rightarrow \min. \quad (9)$$

На каждом этапе поиск осуществляется не только внутри ИИП (k), но и вне его, поэтому вполне возможна такая ситуация, когда $x^k \notin \text{ИИП}(k-1)$. Следовательно, задача (9) распадается на две:

$$\begin{aligned} \sum_{k=1}^{K-1} k \cdot \ell_k &\rightarrow \min, \\ \sum_{k=1}^{K-1} k \cdot \bar{\ell}_k &\rightarrow \min. \end{aligned} \quad (10)$$

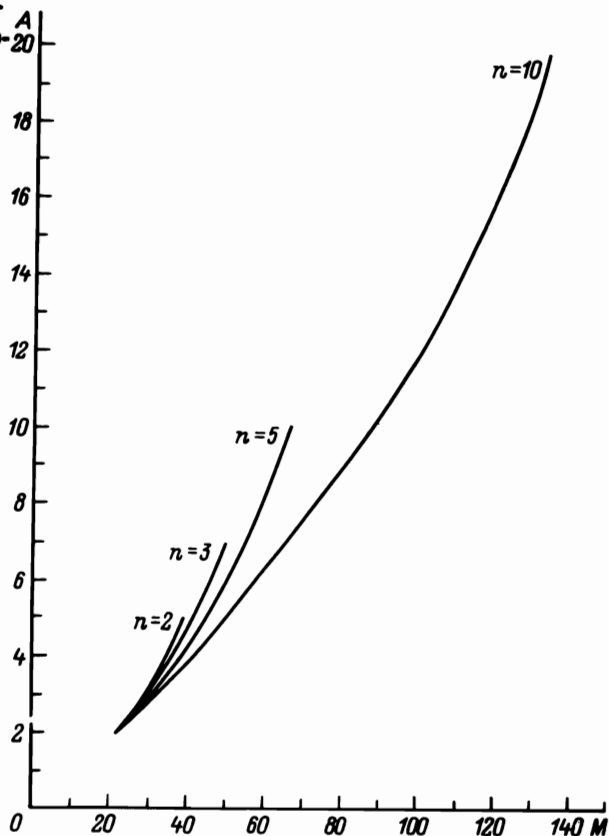
С другой стороны, в целях сокращения затрат машинного времени следует сокращать суммарное число исследуемых точек, а это приводит к задаче $M \cdot K \rightarrow \min$.

Таким образом, задача оптимального выбора параметров A и M свелась к трехкритериальной задаче минимизации $F = \langle f_1, f_2, f_3 \rangle \rightarrow \min$,

где $f_1 = M \cdot K$, а f_2 и f_3 — из формулы (10).

Решение данной трехкритериальной задачи (оптимальное соотношение между А и М) для случаев $n = 2, 3, 5, 10$ и точности $\varepsilon = 1 \cdot 10^{-3}$ показано на рисунке. Кроме того, в целях наглядного геометрического контроля была решена также двухкритериальная задача, в которой второй и третий критерии были свернуты в один с коэффициентами, выравнивающими вклад значений критериев на различных этапах оптимизации. Построенное множество Парето задачи $f_1 = M \cdot K \rightarrow \min; f_2 = \sum_{k=1}^n k \cdot ((1-2q_k) \cdot \bar{e}_k + 2q_k \cdot \bar{e}_k) \rightarrow \min$ достаточно хорошо подтверждает правильность исходных предпосылок. Также были рассмотрены случаи $n = 2, 3, 5, 10$.

В заключение можно отметить, что, исходя из графиков, с увеличением размерности пространства X следует увеличивать величину коэффициента сжатия А, так как влияние параметра n в том виде, в каком он входит в формулу сжатия ИПП(κ) оказывается слишком сильным.



ЛИТЕРАТУРА

Оптимальные соотношения параметров метода ВАСП

1. Туртия С.Б./Об одном способе решения многокритериальных задач//АСУ и приборы автоматизации. 1986, № 77.-С.16-21.
2. Соболев И.М., Статников Р.Б. Выбор оптимальных параметров в задачах со многими критериями. М.: Наука, 1981.
3. Растрингин Л.А., Рипа К.К., Тарасенко Г.С. Адаптация случайного поиска. Рига: Зинатне, 1978.
4. Туртия С.Б. Об ослаблении условий существования и сужения области поиска множества Парето. НТО АН СССР. Л., 1986. 13 с. Рукопись деп. в ВИНТИ 16.06.86. № 4386-В86.