

О ВЕРОЯТНОСТНЫХ МЕТОДАХ ОПТИМИЗАЦИИ НА ДИСКРЕТНЫХ МНОЖЕСТВАХ

С. В. Яковлев

Пусть X дискретное множество, на котором определен функционал $\kappa: X \rightarrow R^1$. Требуется найти

$$x^* = \arg \min_{x \in X} \kappa(x) \quad (1)$$

Рандомизируем задачу (1). Отождествим пространство элементарных событий с X . Пусть \mathcal{B}_X — σ -алгебра подмножеств X (в частности возможно $\mathcal{B}_X = 2^X$), а P_X — вероятностная мера на \mathcal{B}_X . Тогда $\kappa(x)$ есть случайная величина на вероятностном пространстве $\langle X, \mathcal{B}_X, P_X \rangle$.

В данной статье предлагаются методы статистической оптимизации, основанные на использовании свойств случайной величины $\kappa(x)$ для исследования поведения минимизируемого функционала на множестве X и его подмножествах. В дальнейшем, чтобы отличать, идет речь о функционале или о случайной величине, будем использовать соответственно обозначения $\kappa(x)$ и $\kappa(x)$: $x \in Y$, указывая тем самым множество, на котором задана вероятностная мера.

Пусть $\tau = \{Y_j^i\}$, $i = \overline{1, n-1}$, $j = \overline{1, n_i}$ такая система множеств, что

$$Y_j^i \subset \bigcup_{l=1}^{n_{ij}} Y_{\pi_{ij}^l}^{i+1}, \quad Y_1^1 = X, \quad n_1 = 1, \quad \text{card } Y_j^i > \text{card } Y_k^{i+1}, \quad \forall j, k,$$

где $\Pi_{ij} = \{\pi_{ij}^1, \dots, \pi_{ij}^{n_{ij}}\}$ — такие подмножества множества $N_i = \{1, 2, \dots, \sum_{j=1}^{n_i} n_{ij}\}$, что

$$\bigcup_{j=1}^{n_i} \Pi_{ij} = N_i, \quad \Pi_{ij} \cap \Pi_{ik} = \emptyset \quad \forall j, k = \overline{1, n_i}.$$

Для определенности занумеруем, например, множества системы τ так, чтобы

$$Y_j^i \subset \bigcup_{l=\alpha_{j-1}^i+1}^{\alpha_j^i} Y_l^{i+1},$$

где

$$\alpha_0^i = 0, \quad \alpha_k^i = \sum_{j=1}^k n_{ij}, \quad k = \overline{1, n_i}.$$

Ясно, что для построения τ достаточно, например, осуществить разбиение исходного множества X на подмножества, затем каждое из подмножеств снова разбить на подмножества и т. д. Однако заметим, что условие попарного непересечения множеств с одинаковым верхним индексом не является обязательным.

На подмножествах системы τ будем задавать вероятностную меру. Тем самым получим совокупность случайных величин $\varkappa(x)$: $x \in Y \in \tau$. Одной из важнейших характеристик случайной величины является функция ее распределения. Идеи использования функции распределения случайной величины для оптимизации детерминированного функционала рассматривались, например, в работах [1, 2]. Заметим, что в общем случае вид функции распределения $F(v)$ случайной величины $\varkappa(x)$: $x \in Y$ не известен. Однако можно утверждать следующее.

Во-первых, $F(\varkappa(x^*)) = 0$, а для любого $\varepsilon > 0$: $F(\varkappa(x^*) + \varepsilon) > 0$.

Во-вторых, если множество X конечно, а $N(\varepsilon)$ есть множество тех значений $x \in X$, для которых $\varkappa(x) < \varkappa(x^*) + \varepsilon$, то

$$F(\varkappa(x^*) + \varepsilon) = \frac{\text{card } N(\varepsilon)}{\text{card } X}.$$

В-третьих, функция $F(v)$ ступенчатая. Вместе с тем ее можно аппроксимировать непрерывной функцией, воспользовавшись, в частности, разложением $F(v)$ в ряд по нормальным распределениям. При разложении необходимо знать моменты случайной величины. Для этого можно предложить их выборочные оценки. Однако оценки моментов высоких порядков порождают большие погрешности. Поэтому такой подход целесообразен, когда $F(v)$ хорошо приближается 2—3 членами ряда в разложении.

Для некоторых классов функционалов дискретный характер множества X позволяет получать точные значения моментов. В частности, для задачи назначения в работе [3] получены точные среднее и дисперсия, как определенные функции от элементов матрицы стоимости (эти результаты можно перенести на моменты более высоких порядков). Кроме того, сам вид минимизируемого функционала $\varkappa(x)$ иногда позволяет говорить о функции распределения $\varkappa(x)$: $x \in Y$. Так, если $\varkappa(x) = \sum_{i=1}^m c_i x_i$, то при выполнении достаточно общих условий случайная величина $\varkappa(x)$: $x \in Y$ асимптотически ($m \rightarrow \infty$) нормальна. Функцию распределения можно также оценивать с помощью критериев согласия и т. д.

В дальнейшем будем предполагать, что вид функции распределения $F(v)$ известен с точностью до параметров θ и пользоваться обозначением $F(v, \theta)$.

Назовем испытанием генерацию точки x из множества $Y \in \tau$. Для оценки вектора параметров $\theta(Y)$ функции распределения $F(v, \theta(Y))$ случайной величины $\varkappa(x)$: $x \in Y$ произведем серию испытаний (т. е. сформируем выборку $\{\varkappa(x_1), \dots, \varkappa(x_s)\}$, $x_i \in Y$, $i = \overline{1, s}$). Наименьшее выборочное значение при этом

можно рассматривать как приближение к оптимуму функционала $\kappa(x)$. Зная функцию распределения и экстремальное выборочное значение, можно оценивать вероятность генерации точек x , которым соответствуют значения функционала меньшие, чем получены ранее. Тем самым можно осуществлять такой направленный перебор точек множества X , при котором вероятность уменьшения значений функционала растет.

Следующий алгоритм основан на указанных выше соображениях. Предполагается, что для оценки параметров функции распределения и экстремального выборочного значения производится серия из s испытаний, а суммарное число испытаний ограничено и равно S .

Алгоритм 1

Шаг 1. Полагаем $i=j=1$, $S=S-s$, $l_q^p = \alpha_q^p - 1$, $p = \overline{1, n-1}$, $q = \overline{1, n_p}$. Генерируем точки x_1, \dots, x_s из множества X . По выборке $\kappa(x_1), \dots, \kappa(x_s)$ определяем $\kappa(x^0) = \min \{\kappa(x_1), \dots, \kappa(x_s)\}$ и оцениваем параметры $\theta(Y_1^i)$ функции распределения случайной величины $\kappa(x)$: $x \in Y_1^i$. Переходим к шагу 2.

Шаг 2. Если $i \geq n$, переходим к шагу 6; иначе — к шагу 3.

Шаг 3. Полагаем $l_j^j = l_j^j + 1$, $k = l_j^j$. Если $k > \alpha_j^j$, переходим к шагу 6; иначе — к шагу 4.

Шаг 4. Генерируем точки x_1, \dots, x_s из множества Y_k^{i+1} . Вычисляем $\kappa(x^0) = \min \{\kappa(x^0), \kappa(x_1), \dots, \kappa(x_s)\}$ и оцениваем параметры $\theta(Y_k^{i+1})$ функции распределения случайной величины $\kappa(x)$: $x \in Y_k^{i+1}$.

Шаг 5. Определяем такие p и q , для которых значение $F(\kappa(x^0), \theta(Y_q^p))$ наибольшее из всех рассмотренных множеств Y_k^i . Полагаем $i=p$, $j=q$, $S=S-s$ и переходим к шагу 2.

Шаг 6. Генерируем точки x_1, \dots, x_s из множества Y_j^i . Точку x^0 и величину $\kappa(x^0) = \min \{\kappa(x^0), \kappa(x_1), \dots, \kappa(x_s)\}$ принимаем за приближение к оптимуму.

Сделаем некоторые замечания. Алгоритм допускает несложные преобразования в случае если число испытаний s (объем выборки) на каждом шаге различное. Если суммарное число испытаний S не фиксировано, то в качестве S можно взять достаточно большое число. В этом случае, если будет получено уменьшение значения функционала $\kappa(x^0)$ (шаг 6), следует вернуться к шагу 5 и алгоритм продолжит работу.

Согласно алгоритма после каждой серии испытаний необходимо пересчитывать величины $F(\kappa(x^0), \theta(Y_j^i))$ для всех рассмотренных ранее множеств. Но если значение $\kappa(x^0)$ не изменилось, указанные величины также останутся прежними, т. е. вычислять $F(\kappa(x^0), \theta(Y_j^i))$ требуется только для нового сформированного множества.

Обобщением предложенного алгоритма можно считать алгоритмы, в которых на каждом шаге вероятностная мера задается не на Y_j^i , а на всем X . При этом структура вероятностной меры такова, что вероятность генерации точек, „близких“ к рекордной, увеличивается.

При реализации алгоритма 1 вид функции распределения случайных величин $\kappa(x)$: $x \in Y_j^i$ задается априорно или оценивается. Кроме того необходимо получать выборочные оценки параметров функций распределения. Все это порождает определенные погрешности вероятностной модели.

Следующий алгоритм основан на непараметрическом подходе к оценке вероятности улучшений и построен с использованием схемы независимых испытаний (схема Бернулли).

Предположим, что к началу очередной серии испытаний рекордное значение функционала равно $\kappa(x^0)$. Пусть A -случайное событие, состоящее в генерации точки $x \in Y_j^i$, такой что $\kappa(x) < \kappa(x^0)$. Положим, что вероятность события A равна p . Рассмотрим сложное испытание, состоящее в m -кратном повторении простого испытания (генерации точки из Y_j^i). Число λ появлений события A при m -кратном повторении независимых простых испытаний подчиняется биномиальному закону распределения вероятностей

$$P\{\lambda = m\} = \binom{m}{\lambda} p^{\lambda} (1-p)^{m-\lambda}. \quad (2)$$

Итак, величина p — это вероятность улучшений. Если в серии из s испытаний получено l улучшений, то p существенно оценивается отношением l/s . В частности, если улучшение (событие A) произошло на k -м испытании, то $p = 1/k$. Зная оценку p , а также значения λ и $P\{\lambda = m\}$, из выражения (2) несложно оценить длину серии испытаний до получения λ улучшений (успехов).

Алгоритм 2

- Шаг 1.* Полагаем $i=j=l=1$, $v_0=v^*=10^{16}$. Задаем (или оцениваем) длину серии испытаний s и величину ожидаемых успешных испытаний λ .
- Шаг 2.* Формируем множество Y_j^i и задаем вероятностное распределение на нем. Полагаем $k=0$, $m=0$. Переходим к шагу 3.
- Шаг 3.* Полагаем $k=k+1$. Если $k>s$, переходим к шагу 8; иначе — к шагу 4.
- Шаг 4.* Генерируем точку $x \in Y_j^i$ и вычисляем $\kappa(x)$. Если $\kappa(x) < v_0$, полагаем $x^0=x$, $v_0=\kappa(x^0)$. Переходим к шагу 5.
- Шаг 5.* Если $\kappa(x) < v^*$, полагаем $m=m+1$. Если $m < \lambda$, переходим к шагу 3; иначе — полагаем $v^*=v_0$ и переходим к шагу 6.
- Шаг 6.* Полагаем $j=l$, $l=\alpha_{j+1}^i+1$ и переходим к шагу 7.
- Шаг 7.* Полагаем $i=i+1$. Если $i < n$, переходим к шагу 2; иначе — к шагу 9.
- Шаг 8.* Если $l < \alpha_j^i$, полагаем $l=l+1$ и переходим к шагу 2; иначе — переходим к шагу 6.
- Шаг 9.* Точку x^0 и величину $\kappa(x^0)$ принимаем за приближение к оптимуму.

В рамках приведенного алгоритма можно варьировать числом ожидаемых успешных испытаний, уменьшая λ при уменьшении $\text{card } Y_i^j$. Тот факт, что на начальных этапах число λ целесообразно брать большим, объясняется необходимостью более тщательного „просмотра” множества Y_j^i для определения перспективного направления дальнейшего поиска.

Заметим, что как в первом, так и во втором алгоритмах важное значение имеет правило выбора нового множества Y_i^j . Естественно это множество выбирать так, чтобы рекордная точка x^0 ему принадлежала. Если такое множество уже рассмотрено, выбирается множество, которому принадлежит ближайшая (по значению функционала) точка и т. д. В приведенных алгоритмах указанный шаг опущен в целях простоты изложения, но такая процедура легко реализуется путем соответствующей перенумерации множеств системы τ .

В задачах оптимизации важное значение имеет поведение функционала $\kappa(x)$ в окрестности глобального экстремума. В терминах вероятностной модели это соответствует поведению функции распределения $F(v)$ случайной величины $\kappa(x)$: $x \in Y$ на „хвостах”. Поскольку функционал $\kappa(x)$ ограничен на Y снизу, то положив

$$\eta_Y = \inf_{x \in Y} \kappa(x),$$

можно утверждать, что

$$F(\eta_Y) = 0, \quad F(\eta_Y + \delta) \neq 0 \quad \forall \delta > 0. \quad (3)$$

Зададимся числом $\varepsilon > 0$. Поставим функции распределения $F(v)$, удовлетворяющей условию (3), в соответствие функцию распределения $F^*(v)$ такую, что для некоторого числа η_Y^* имеет место:

$$F^*(\eta_Y^*) = 0, \quad F^*(\eta_Y^* + \delta) \neq 0 \quad \forall \delta > 0, \quad |\eta_Y - \eta_Y^*| < \varepsilon.$$

Тогда параметр η_Y^* является статистической оценкой оптимума функционала $\kappa(x)$ на Y . В качестве $F^*(v)$ могут выступать предельное распределение случайной величины $\kappa(x)$: $x \in Y$, а также предельное распределение ее экстремальных значений. В последнем случае точный вид исходного распределения $F(v)$ нас не интересует. Важно, чтобы оно удовлетворяло постулату устойчивости [4].

При статистической оценке оптимума можно использовать и непараметрический подход, в основу которого положены элементы теории порядковых статистик. Рассмотрим независимые реализации $\kappa(x_1), \dots, \kappa(x_s)$ случайной величины $\kappa(x)$: $x \in Y$. Обозначим через $\eta_{(1)} \leq \dots \leq \eta_{(s)}$ порядковые статистики, соответствующие данной выборке. Выберем k крайних порядковых статистик и рассмотрим величину

$$\hat{\eta}(Y) = \sum_{i=1}^k a_i \eta_{(i)}.$$

Правило выбора k и значения коэффициентов a_i , $i = \overline{1, k}$ приведены в [2, 5]. Заметим, что величина $\hat{\eta}(Y)$ состоятельно оценивает η_Y^* . Возможность статистического оценивания оптимума функционала $\kappa(x)$ на множествах Y_j^i позволяет предложить группу алгоритмов, построенных аналогично методу ветвей и границ.

Алгоритм 3.

Шаг 1. Полагаем $i=j=1$, $v_0=10^{16}$, $l_p=1$, $p=\overline{0, n}$. Задаемся допустимой погрешностью оценки $\bar{\varepsilon}>0$.

Шаг 2. Генерируем точки x_1, \dots, x_s из множества Y_j^i . Полагаем $v_0 = \min \{v_0, \kappa(x_1), \dots, \kappa(x_s)\}$. Вычисляем статистическую оценку оптимума $\hat{\eta}(Y_j^i)$. Если $\hat{\eta}(Y_j^i) > \kappa(x^0) - \bar{\varepsilon}$, переходим к шагу 4; иначе — к шагу 3.

Шаг 3. Полагаем $j=\alpha_{i-1}^i$, $i=i+1$. Если $i>n$, переходим к шагу 5; иначе — к шагу 4.

Шаг 4. Полагаем $j=j+1$, $l_i=j$. Если $j \leq \alpha_{i-1}^i$, переходим к шагу 2; иначе — к шагу 5.

Шаг 5. Полагаем $i=i-1$, $j=l_i$. Если $i \leq 1$, переходим к шагу 7; иначе — к шагу 6.

Шаг 6. Если $\hat{\eta}(Y_j^i) > \kappa(x^0) - \bar{\varepsilon}$, переходим к шагу 5; иначе — к шагу 4.

Шаг 7. Точку x^0 и величину $\kappa(x^0)$ принимаем за приближение к оптимуму.

Как видно из приведенного алгоритма, он состоит из прямого хода, когда вычисляются статистические оценки, и обратного хода, когда эти оценки используются для отсечений. Прямой ход определяется шагами 2—4, а обратный — шагами 5, 6. Погрешность $\bar{\varepsilon}$ оценки может задаваться априорно, что будет соответствовать некоторому уровню вероятности, либо вычисляться с учетом функции распределения случайной величины $\hat{\eta}(Y_j^i)$.

Заметим, что необходимость многократной статистической оценки оптимума на подмножествах Y_j^i увеличивает вероятность утери множества Y_j^i , которому принадлежит глобальный экстремум. Однако можно использовать дополнительные средства в процессе поиска.

Имеется в виду следующее. Если $Y_2 \subset Y_1$, то

$$\min_{x \in Y_1} \kappa(x) \leq \min_{x \in Y_2} \kappa(x).$$

Таким образом вместо непосредственной оценки $\hat{\eta}(Y_1)$ и $\hat{\eta}(Y_2)$ можно проверить гипотезу $H_0: \hat{\eta}(Y_2) - \hat{\eta}(Y_1) \leq \bar{\varepsilon}$ при альтернативе $H_1: \hat{\eta}(Y_2) - \hat{\eta}(Y_1) > \bar{\varepsilon}$. Другими словами, проверяется гипотеза, принадлежит ли искомое приближение к оптимуму подмножеству множества, содержащего это приближение.

Алгоритм 4.

Шаг 1. Полагаем $i=j=1$, $v_0=10^{16}$. Задаемся погрешностью $\bar{\varepsilon}>0$.

Шаг 2. Генерируем точки $x_1, \dots, x_s \in Y_1^i$. Вычисляем $\eta^0 = \hat{\eta}(Y_1^i)$, $v_0 = \min \{v_0, \kappa(x_1), \dots, \kappa(x_s)\}$.

Шаг 3. Полагаем $j=\alpha_{i-1}^i$, $i=i+1$. Если $i>n$, переходим к шагу 8; иначе — к шагу 4.

Шаг 4. Генерируем точки $x_1, \dots, x_s \in Y_j^i$. Вычисляем статистическую оценку оптимума $\hat{\eta}(Y_j^i)$ и $v_0 = \min \{v_0, \kappa(x_1), \dots, \kappa(x_s)\}$.

Шаг 5. Проверяем гипотезу $H_0: \hat{\eta}(Y_j^i) - \eta^0 \leq \bar{\epsilon}$ при альтернативе $H_1: \hat{\eta}(Y_j^i) - \eta^0 > \bar{\epsilon}$. Если гипотеза принимается, переходим к шагу 3; иначе — к шагу 6.

Шаг 6. Полагаем $j=j+1, l_i=j$. Если $j \leq \alpha_{l_i-1}^i$, переходим к шагу 4; иначе — к шагу 7.

Шаг 7. Полагаем $i=i-1, j=l_i$. Если $i \geq 1$, переходим к шагу 4; иначе — к шагу 8.

Шаг 8. Точку x^0 и величину $\kappa(x^0)$ принимаем за начальное приближение к оптимуму. Если $i > n$, дальнейший поиск осуществляется на множестве Y_{n-1}^n .

Заметим, что при отсутствии погрешностей вычисления, если $Y \subset \bigcup_{i=1}^k Y_i$,

то существует такое $j = \overline{1, k}$, что $\eta_Y = \eta_{Y_j}$. Однако вероятностный характер оценок $\hat{\eta}(Y_i)$ может привести к тому, что для всех $j = \overline{1, k}$: $\hat{\eta}(Y_i) - \hat{\eta}(Y) > \bar{\epsilon}$. Поэтому на шаге 7 алгоритма по-существу осуществляется возврат к уже рассмотренному ранее множеству и пересчет оценки $\hat{\eta}(Y_j^i)$.

Сравнение статистических оценок оптимумов может осуществляться не только относительно исходного множества X , а и относительно множества, полученного на предыдущем этапе. В этом случае проверяется гипотеза $H_0: \hat{\eta}(Y_j^i) - \hat{\eta}(Y_{l_i-1}^i) > \bar{\epsilon}$.

Во всех приведенных в статье алгоритмах общим является использование статистических свойств функционала $\kappa(x)$ на подмножествах множества X для определения направления поиска оптимума. Эти алгоритмы могут быть так или иначе модифицированы применительно к конкретно решаемому классу задач. В частности, особенности применения алгоритмов в задачах размещения, квадратичного назначения, компоновки, балансировки и т. д. рассмотрены в [6—10]. Обширный вычислительный эксперимент описан в [2].

В заключение отметим, что класс предложенных алгоритмов объединен общим названием — метод последовательной статистической оптимизации [2]. Начало разработкам положено членом-корреспондентом АН УССР, профессором Ю. Г. Стояном [11]. Теоретические аспекты обоснования метода нашли отражение в [2, 12]. Более подробную библиографию также можно найти в [2].

Автор выражает благодарность Ю. Г. Стояну за постоянное внимание и помощь в работе.

ХАРЬКОВСКИЙ ИНСТИТУТ РАДИОЭЛЕКТРОНИКИ
КАФЕДРА ПРИКЛАДНОЙ МАТЕМАТИКИ
Г. ХАРЬКОВ, ПРОСП. ЛЕНИНА, 14.
СССР, 310141.

Литература

- [1] Моцкус, И. Б., Многоэкстремальные задачи в проектировании (Статистические решения. Усиление локальных методов. Эвристические способности человека). — М.: Наука, 1967. — 215 с.

- [2] Стоян, Ю. Г., Яковлев, С. В., Математические модели и оптимизационные методы геометрического проектирования. — Киев: Наукова думка, 1986. — 276 с.
- [3] Ходзинский, А. Н., Статистические свойства критерия и их использование в алгоритмах решения задач комбинаторной оптимизации. — В кн.: Применения случайного поиска, Кемерово, 1984, с. 41—42.
- [4] Гумбель Э., Статистика экстремальных значений. — М.: Мир, 1965. — 450 с.
- [5] Математическая теория планирования эксперимента (Под ред. С. М. Ермакова). — М.: Наука, 1983. — 392 с.
- [6] Стоян, Ю. Г., Гиль Н. И., Методы и алгоритмы размещения плоских геометрических объектов. — Киев: Наукова думка, 1976. — 247 с.
- [7] Стоян, Ю. Г., Туранов Н. Т., Алгоритм размещения плоских фигур с наименьшей длиной связывающей сети. — Управляющие системы, 1970, вып. 4/5, с. 97—107.
- [8] Стоян, Ю. Г., Соколовский, В. З., Решение некоторых многоэкстремальных задач методом сужающихся окрестностей. — Киев: Наукова думка, 1981. — 205 с.
- [9] Кухаренок, М. А., Совместное решение задач размещения элементов и внешних контактных площадок при проектировании ЦВА. — В кн.: Прикладные методы кибернетики. — Киев: ИК АН УССР, 1984, с. 72—78.
- [10] Стоян, Ю. Г., Соколовский, В. З., Яковлев, С. В., Метод уравнивания вращающихся дискретно распределенных масс. — Энергомашвиностроение, 1982, № 2, с. 4—5.
- [11] Стоян, Ю. Г., Об одном способе поиска наилучшего решения для одного класса многоэкстремальных задач. — Управляемые системы, 1969, вып. 3.
- [12] Стоян, Ю. Г., Яковлев, С. В., Исследование сходимости и эффективности метода сужающихся окрестностей. — Харьков, 1981. — 43 с. (Препринт АН УССР.) Ин-т пробл. машиностроения; № 168.)

(Received March 21, 1986)