

ОБЩИЙ ПЛАН РЕШЕНИЯ ЗАДАЧ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Материал подготовил ассистент А. Тонис

Рассматривается конечномерная задача на условный экстремум (максимум или минимум):

$$\begin{aligned} f(x) &\rightarrow \text{extr} \\ \text{s. t. } x &\in M = \{x \mid g_i(x) \geq 0, \quad i = 1, \dots, m\}, \end{aligned} \quad (1)$$

где $x = (x_1, \dots, x_n)^\top$, а функции f, g_1, \dots, g_m дважды непрерывно дифференцируемы в некоторой открытой области в \mathbf{R}^n , но никаких предположений относительно вогнутости или выпуклости на них не накладывается. Предлагается следующий план решения задачи (1):

- 1) Выписать функцию Лагранжа

$$L(x, \lambda) = f(x) + \lambda g(x), \quad (2)$$

где $\lambda = (\lambda_1, \dots, \lambda_m)$, $g = (g_1, \dots, g_m)^\top$. При этом, если рассматривается задача на максимум, то на λ накладывается ограничение $\lambda_i \geq 0$, а если задача на минимум, то $\lambda_i \leq 0$.

- 2) Выписать условия стационарности
- L
- по
- x
- , т. е. систему из
- n
- уравнений вида

$$\frac{\partial L}{\partial x_i} = 0, \quad i = 1, \dots, n. \quad (3)$$

Попытаться преобразовать (3) к более удобному для использования виду, например, выразить x через λ , если это возможно.

- 3) Подключить
- m
- условий дополняющей нежесткости вида

$$\lambda_i g_i(x) = 0 \quad (4)$$

и, таким образом, получить систему из $m+n$ уравнений (3)–(4) с $m+n$ неизвестными (x, λ) , которую надо будет решать. На практике для решения системы (3)–(4) придется рассмотреть все возможные комбинации активных и неактивных ограничений или, что почти то же самое, все варианты равенства/неравенства нулю множителей Лагранжа (например, $\lambda_1 = 0, \lambda_2 \neq 0, \dots$). Всех таких комбинаций, вообще говоря, много (2^m), но исходя из различных соображений (геометрических, здравого смысла, несовместности некоторых уравнений, — например, если слишком много ограничений предполагаются активными), часто можно сузить объем перебора.

- 4) Итак, получены все решения
- (x, λ)
- системы (3)–(4). Это, в самом первом приближении, точки, подозрительные на экстремум. Для дальнейшей их проверки и отсева надо убедиться, что для каждой подозрительной точки
- \bar{x}
- и соответствующих
- λ
- помимо равенств (3)–(4) выполнены следующие неравенства:

- а) Все множители Лагранжа имеют “правильный” знак (для функции Лагранжа вида (2) $\lambda_i \geq 0$ / $\lambda_i \leq 0$ в случае задачи на максимум/минимум). Если есть как положительные, так и отрицательные λ_i , то точка \bar{x} заведомо не является экстремумом.
- б) Точка \bar{x} действительно удовлетворяет всем ограничениям $g_i(x) \leq 0$. Реально, конечно, придется проверять лишь неактивные ограничения.

- 5) Все, что перечислено выше, — это
- условия первого порядка*
- , использующие линеаризацию задачи в окрестности подозрительной точки, когда учитываются лишь значения функций и их первые производные в этой точке. Условия первого порядка являются необходимыми, но не достаточными, т. е. позволяют сузить круг подозреваемых точек, но не доказывают их оптимальность. Для окончательного определения экстремумов требуется более детальная информация, касающаяся
- вторых*
- производных. А именно, надо исследовать на положительную или отрицательную определенность квадратичную форму

$$H(z) = z^\top L''(\bar{x})z, \quad (5)$$

где L'' — гессиан (матрица вторых производных) функции L . Важный момент состоит в том, что квадратичная форма H задана не на всем \mathbf{R}^n , а на более узком пространстве T — касательном подпространстве к множеству, задаваемому *связывающими* ограничениями (в отличие от активного, связывающего ограничения для точки \bar{x} — это когда не только $g_i(\bar{x}) = 0$, но и $\lambda_i > 0 / < 0$ для задачи на \max / \min).

Как же найти T ? Надо воспользоваться тем, что T есть ортогональное дополнение к линейной оболочке градиентов связывающих ограничений, т. е. все касательные векторы $z \in T$ (и только они) удовлетворяют системе линейных уравнений

$$g'_i(\bar{x})z = 0, \quad i \in I, \quad (6)$$

где I — множество связывающих ограничений. Таким образом, надо найти пространство решений системы (6) и исследовать на нем квадратичную форму H . Если $\lambda \geq 0$ и H (строго) положительно определенная, то \bar{x} — локальный минимум; если $\lambda \leq 0$ и H (строго) отрицательно определенная, то \bar{x} — локальный максимум.

Выше было приведено *достаточное* условие второго порядка, применяемое для выявления экстремумов. Чтобы убедиться в том, что точка \bar{x} *не* является экстремумом, надо проверить, что не выполняется *необходимое* условие второго порядка. А именно, надо рассмотреть $H(z)$ на пространстве \tilde{T} — касательном подпространстве к множеству, задаваемому *активными* ограничениями (ясно, что $\tilde{T} \subseteq T$). Если, согласно условиям первого порядка, точка \bar{x} подозрительна на минимум (максимум), и в то же время $\exists z \in \tilde{T} : H(z) < 0$ (> 0), то \bar{x} — не экстремум.

Бывают “пограничные” случаи, не подпадающие под приведенные выше условия (например, функция $f(x) = x^3$ при $x = 0$). Это значит, что не получилось идентифицировать экстремум с помощью условий первого и второго порядка и требуется более тонкий анализ.

Как проверить H на положительную (отрицательную) определенность? Если $\dim T = 1$, то достаточно выбрать любой ненулевой вектор $z \in T$ и вычислить $H(z)$ по формуле (5). При большей размерности T придется построить в T ортогональный базис, найти матрицу квадратичной формы H в этом базисе и применить критерий Сильвестра. Необходимо иметь в виду, что если точка x — *вершина*, т. е. имеется n активных в ней ограничений с линейно независимыми градиентами и, кроме того, все эти ограничения являются связывающими для рассматриваемой задачи, то $\dim T = 0$, и поэтому достаточное условие второго порядка выполняется автоматически, так что проверять его не надо.

Для проверки условий второго порядка иногда необязательно знать T или \tilde{T} . Например, если квадратичная форма (5) является положительно определенной на всем \mathbf{R}^n , то она по-прежнему обладает тем же свойством в ограничении на любое подпространство.

- 6) Метод, описанный выше, позволяет устанавливать лишь *локальные* экстремумы задачи (1). Для того, чтобы получить *глобальный* экстремум, надо вычислить значения целевой функции f во всех локальных экстремумах и выбрать из них наибольшее/наименьшее. Более того, если нас интересует исключительно глобальный экстремум, оказывается достаточным просто сравнить значения f вообще во всех точках, подозрительных на экстремум (и удовлетворяющих всем ограничениям!), таким образом освободив себя от трудоемкой проверки условий второго порядка. Все это, однако, работает только в том случае, когда глобальный экстремум *существует*. Он может не существовать по одной из двух причин: f неограничена (с соответствующей стороны) на M , либо f ограничена, но \inf или \sup не достигается ни при каком конечном x из-за неограниченности самого M . Существование обоих глобальных экстремумов гарантировано во всех случаях, когда M — замкнутое ограниченное множество, а функция f непрерывна на нем.

Примечание. Условия первого порядка основаны на линеаризации всех функций в окрестности интересующей нас точки x . Поэтому наш метод гарантирует обнаружение всех локальных экстремумов лишь в том случае, когда линеаризация качественно не меняет вид допустимой области M в окрестности x . На рис. 1 изображена ситуация, когда это не так. Видно, что если линеаризовать границы области в окрестности “клюва”, то она станет неограниченной и экстремум достигаться не будет. Отсутствие такой “патологии”, как на рис. 1, гарантируется условием регулярности, которое требует, чтобы градиенты к активным ограничениям были линейно независимы.

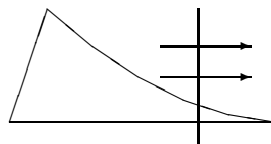


Рис. 1. Невыполнение условия регулярности