ГЛАВА Функции нескольких переменных (продолжение)

§ Численные методы поиска экстремума

В предыдущем параграфе были сформулированы необходимые и достаточные условия безусловного локального экстремума. Однако, их далеко не всегда удается использовать на практике. Чтобы проверить необходимые условия экстремума, надо решить систему алгебраических уравнений, что редко удается сделать аналитически. Целевая функция, экстремум которой ищется, может оказаться сложной, не дифференцируемой, или вообще, не заданной в виде конкретной формулы. В таких случаях целесообразно задействовать численные методы поиска экстремума.

1. Оптимизация функций одной переменной. Пусть требуется найти наибольшее и наименьшее значения функции $f(x)$ на отрезке $[a,b]$. Разобьем отрезок $[a,b]$ на n произвольных частей:

$$a = x_0 < x_1 < \ldots < x_n = b.$$

Вычислим $y_i = f(x_i)$, $i = 0, \ldots, n$. Тогда можно предположить, что

$$\max_{[a,b]} f(x) \approx \max\{y_0, \ldots, y_n\}, \quad \min_{[a,b]} f(x) \approx \min\{y_0, \ldots, y_n\}.$$

При этом возникают проблемы: Как лучше выбрать значение n? Как лучше расположить точки дробления $x_1, \ldots, x_n$? Как оценить погрешность найденных экстремальных значений? Как правило, для решения этих проблем приходится задействовать "физику" процесса.

2. Оптимизация функций нескольких переменных. Пусть теперь требуется найти глобальные экстремумы у функции $f(x_1, \ldots, x_n)$ в некоторой области $D \in R^n$. Простое построение сетки узлов на области D (по аналогии со случаем функций одной переменной) приведет к большим вычислительным затратам. Поэтому обычно используют один из следующих подходов (и их различные модификации).

а) Метод покоординатного спуска.

Выберем произвольную начальную точку $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)})^T \in D$. Решим одномерную оптимизационную задачу:

$$f(x_1, x_2^{(0)}, \ldots, x_n^{(0)}) \to \max_{x_1} \left(\min_{x_1}\right).$$

Пусть $x_1^{(1)}$ — решение этой задачи. Теперь оптимизируем значение второй переменной:

$$f(x_1^{(1)}, x_2, x_3^{(0)}, \ldots, x_n^{(0)}) \to \max_{x_2} \left(\min_{x_2}\right).$$

Находим решение $x_2^{(1)}$ этой задачи, и т.д. В результате за n шагов придем в некоторую точку $\mathbf{x}^{(1)} = (x_1^{(1)}, \ldots, x_n^{(1)})^T \in D$. Далее все повторяем заново. Получаем последовательность точек $\{\mathbf{x}^{(k)}\}_{k=0,1,\ldots}$, которая будет приближаться к искомому экстремуму. Остановившись на некотором значении k, найдем приближенное значение экстремума.

б) Метод градиентного спуска.

Снова выберем произвольную начальную точку $\mathbf{x}^{(0)} \in D$. Зададим некоторый шаг $h > 0$. Положим

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + h\vec{\nabla} f(\mathbf{x}^{(0)}),$$

если решаем задачу на поиск максимума, и

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - h\vec{\nabla} f(\mathbf{x}^{(0)}),$$

если решаем задачу на поиск минимума. И т.д. В результате также приходим к последовательности точек $\{\mathbf{x}^{(k)}\}_{k=0,1,\dots}$, приближающейся к экстремуму.

в) Метод случайного спуска.

Выбираем произвольную начальную точку $\mathbf{x}^{(0)} \in D$. Зададим некоторый шаг $h > 0$. На каждой итерации алгоритма выбираем произвольное (случайное) направление $\vec{l}$. Предположим, что решается задача на поиск максимума (для минимума все аналогично). Тогда положим

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + h\vec{l},$$

если $f(\mathbf{x}^{(0)} + h\vec{l}) > f(\mathbf{x}^{(0)})$, и

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)},$$

если $f(\mathbf{x}^{(0)} + h\vec{l}) \leq f(\mathbf{x}^{(0)})$. И т.д. Перебирая случайно на каждой итерации алгоритма различные направления, получим последовательность точек $\{\mathbf{x}^{(k)}\}_{k=0,1,\dots}$, приближающуюся к экстремуму.

Если сравнивать между собой описанные три подхода, то метод покоординатного спуска наиболее простой, но наиболее медленно сходящийся. Метод градиентного спуска быстро сходящийся, но более сложный (он требует на каждом шаге вычисления градиента функции). Метод случайного спуска представляет собой нечто среднее между первыми двумя подходами. Есть различные модификации описанных подходов, ориентированные на решение тех или иных специфических задач.

При реализации описанных подходов приходится сталкиваться со следующими стандартными проблемами:

1) Проблема многоэкстремальности. Экстремумов у функции может быть много, и где они расположены, мы не знаем. Применяя какой-то метод, мы найдем какой-то из экстремумов, но не факт, что глобальный. Чтобы повысить вероятность обнаружения глобального экстремума, часто проводят спуск, начиная с различных начальных точек $\mathbf{x}^{(0)} \in D$.

2) Проблема выбора шага. В методах градиентного и случайного спуска требуется задать какой-то шаг h. Если его задать большим, то мы можем "перепрыгнуть" через экстремум. Если задать малым, то движение в сторону экстремума будет долгим. Обычно сначала выбирают большой шаг, а в процессе движения к экстремуму его постепенно уменьшают.

3) В градиентном методе предполагается вычисление на каждой итерации алгоритма градиента функции. Это может оказаться проблематично, если функция сложная, или неизвестная в виде формулы. Есть различные методы приближенной оценки градиента. Если функция недифференцируемая, то есть специальные методы негладкой оптимизации.

4) Все описанные подходы хорошо работают, если искомый экстремум располагается внутри области D. Если же экстремум располагается на границе или вблизи ее, возникает множество сложностей. В этом случае используют специальные алгоритмы, составляющие основу такого раздела теории оптимизации, как "Математическое программирование".

Подробнее проблема поиска экстремума функции в нетривиальных случаях исследуется в других дисциплинах ("Численные методы", "Линейное и динамическое программирование", "Методы оптимизации", "Теория оптимальных решений", "Методы негладкого анализа", и т.д.)

§ Теорема о неявной функции

Пусть имеется уравнение
$$F(x,y)=0. \tag{1}$$

Определение. *Будем говорить, что уравнение (1) задает функцию $y=y(x)$ неявно, если $F(x,y(x))\equiv 0$.*

Пример. Пусть уравнение (1) имеет вид
$$x^2+y^2=1. \tag{2}$$
Оно неявно задает функции $y=\sqrt{1-x^2}$ и $y=-\sqrt{1-x^2}$. Заметим, что через каждую точку окружности (2) проходит только одна из этих кривых, за исключением двух точек $(-1,0)$ и $(1,0)$ (в этих двух точках явная кривая определена неоднозначно).

Теорема (о неявной функции). *Пусть точка $M_0=(x_0,y_0)$ удовлетворяет уравнению (1), функция $F(x,y)$ непрерывна в окрестности этой точки и имеет там непрерывную частную производную $\frac{\partial F}{\partial y}$, причем $\left.\frac{\partial F}{\partial y}\right|_{M_0}\neq 0$. Тогда можно указать такое $\delta>0$, что найдется единственная непрерывная функция $y=y(x)$, заданная на интервале $(x_0-\delta,x_0+\delta)$ и удовлетворяющая условиям:*

1) $F(x,y(x))\equiv 0$ при $x\in(x_0-\delta,x_0+\delta)$,

2) $y(x_0)=y_0$.

Если дополнительно предположить, что функция $F(x,y)$ имеет в окрестности точки M_0 непрерывную частную производную $\frac{\partial F}{\partial x}$, то тогда указанная функция $y(x)$ будет дифференцируемой на интервале $(x_0-\delta,x_0+\delta)$, и
$$y'(x)=-\frac{F'_x}{F'_y}.$$

Доказательство. По условию теоремы $\left.\frac{\partial F}{\partial y}\right|_{M_0}\neq 0$. Пусть для определенности $\left.\frac{\partial F}{\partial y}\right|_{M_0}>0$ (в противном случае все аналогично). Функция $\frac{\partial F}{\partial y}$ — непрерывна в окрестности точки M_0, значит, можно найти такое $\delta_1>0$, что $\frac{\partial F}{\partial y}>0$ при $(x,y)\in\Omega$, где
$$\Omega=\{(x,y):\ x\in(x_0-\delta_1,x_0+\delta_1),\ y\in(y_0-\delta_1,y_0+\delta_1)\}.$$
Тогда функция $F(x_0,y)$ строго возрастает по y на промежутке $[y_0-\delta_1,y_0+\delta_1]$. Поскольку $F(x_0,y_0)=0$, то получаем
$$F(x_0,y_0-\delta_1)<0,\quad F(x_0,y_0+\delta_1)>0.$$
Функция $F(x,y)$ — непрерывна в окрестности точки M_0, значит, можно найти такое $\delta\in(0,\delta_1]$, что
$$F(x,y_0-\delta_1)<0,\quad F(x,y_0+\delta_1)>0$$
при $x\in(x_0-\delta,x_0+\delta)$ (см. рисунок 1).

Тогда для $\forall\bar{x}\in(x_0-\delta,x_0+\delta)$ имеем
$$F(\bar{x},y_0-\delta)<0,\quad F(\bar{x},y_0+\delta)>0,$$
что по теореме Коши — Больцано влечет: $\exists\bar{y}\in[y_0-\delta_1,y_0+\delta_1]$: $F(\bar{x},\bar{y})=0$, причем, в силу строгого возрастания функции $F(\bar{x},y)$ по y на промежутке $[y_0-\delta_1,y_0+\delta_1]$,

получаем, что значение $\bar{y}$ будет единственным. Таким образом, построили непрерывную функцию $y(x)$ на интервале $(x_0 - \delta, x_0 + \delta)$, удовлетворяющую условиям 1) и 2) теоремы.

Докажем вторую часть теоремы. Выберем произвольную точку $\bar{x} \in (x_0 - \delta, x_0 + \delta)$ и зададим приращение этого аргумента: Δx (считаем, что $(\bar{x} + \Delta x) \in (x_0 - \delta, x_0 + \delta)$). Положим $y(\bar{x}) = \bar{y}$, $y(\bar{x} + \Delta x) = \bar{y} + \Delta y$.

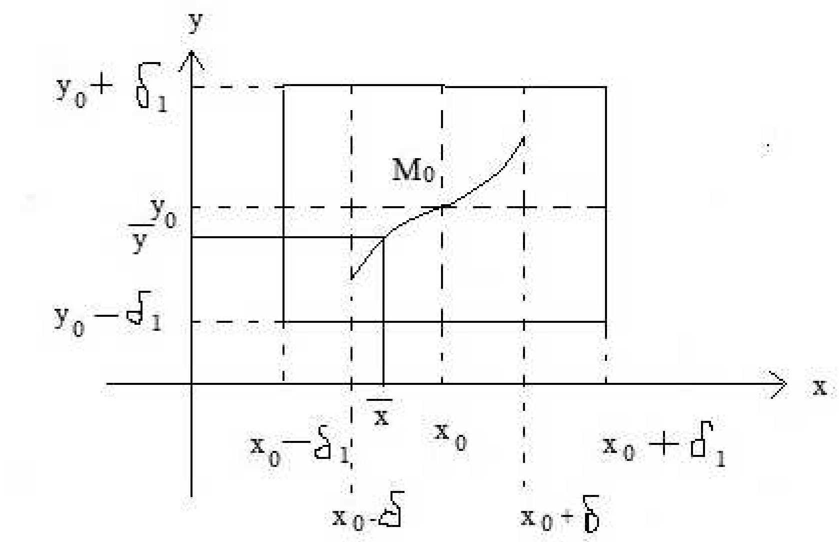

Рис. 1.

Обе точки $(\bar{x}, \bar{y})$ и $(\bar{x} + \Delta x, \bar{y} + \Delta y)$ принадлежат найденной кривой $y = y(x)$, т.е.

$$F(\bar{x}, \bar{y}) = 0, \quad F(\bar{x} + \Delta x, \bar{y} + \Delta y) = 0.$$

Тогда, учитывая дифференцируемость функции $F(x, y)$, получаем

$$0 = \Delta F = F(\bar{x} + \Delta x, \bar{y} + \Delta y) - F(\bar{x}, \bar{y}) = F'_x(\bar{x}, \bar{y})\Delta x + F'_y(\bar{x}, \bar{y})\Delta y + o(\rho),$$

где $\rho = \sqrt{(\Delta x)^2 + (\Delta y)^2}$. Отсюда

$$0 = F'_x(\bar{x}, \bar{y}) + F'_y(\bar{x}, \bar{y})\frac{\Delta y}{\Delta x} + \frac{o(\rho)}{\Delta x}.$$

Переходя к пределу при $\Delta x \to 0$, установим требуемое: $y'(\bar{x}) = -F'_x(\bar{x}, \bar{y})/F'_y(\bar{x}, \bar{y})$. Теорема доказана.

Для вычисления производной функции, заданной неявно, можно использовать обычные правила дифференцирования сложных функций. Пусть функция $y(x)$ задается неявно уравнением (1). Тогда получаем тождество

$$F(x, y(x)) \equiv 0.$$

Продифференцируем его

$$F'_x + F'_y y' \equiv 0.$$

Отсюда получаем формулу для вычисления первой производной $y'(x) = -F'_x/F'_y$. Если продифференцировать это тождество еще раз:

$$\left(F''_{x^2} + F''_{xy}y'\right) + \left[\left(F''_{yx} + F''_{y^2}y'\right)y' + F'_y y''\right] \equiv 0,$$

то получим выражение для второй производной $y''(x)$, и т.д.

Пример. Пусть

$$x^2 + y^2 = 1.$$

Тогда

$$2x + 2yy' = 0,$$

т.е. $y'(x) = -x/y$. Далее

$$1 + y'^2 + yy'' = 0,$$

откуда $y'' = -(1 + y'^2)/y = -(x^2 + y^2)/y^3$, и т.д.

Аналогично можно ввести понятие неявной функции нескольких переменных.

Пусть имеется уравнение

$$F(x_1, \dots, x_n, y) = 0. \tag{3}$$

Определение. *Говорят, что уравнение (3) задает функцию* $y = y(x_1, \dots, x_n)$ *неявно, если* $F(x_1, \dots, x_n, y(x_1, \dots, x_n)) \equiv 0$.

Теорема (о неявной функции нескольких переменных). *Пусть точка* $M_0 = (\mathbf{x}_0, y_0)$*, где* $\mathbf{x}_0 \in R^n$*, удовлетворяет уравнению (3), функция* $F(x_1, \dots, x_n, y)$ *непрерывна в окрестности этой точки и имеет там непрерывную частную производную* $\frac{\partial F}{\partial y}$*, причем* $\left.\frac{\partial F}{\partial y}\right|_{M_0} \neq 0$*. Тогда можно указать такое* $\delta > 0$*, что найдется единственная непрерывная функция* $y = y(x_1, \dots, x_n)$*, заданная в* δ*-окрестности точки* $\mathbf{x}_0$ *и удовлетворяющая там условиям:*

1) $F(x_1, \dots, x_n, y(x_1, \dots, x_n)) \equiv 0$,

2) $y(\mathbf{x}_0) = y_0$.

Если дополнительно предположить, что функция $F(x_1, \dots, x_n, y)$ *имеет в окрестности точки* M_0 *непрерывные частные производные* $\frac{\partial F}{\partial x_i}$, $i = 1, \dots, n$*, то тогда указанная функция* $y(x_1, \dots, x_n)$ *будет иметь частные производные в построенной окрестности точки* $\mathbf{x}_0$*, и*

$$\frac{\partial y}{\partial x_i} = -\frac{F'_{x_i}}{F'_y}, \quad i = 1, \dots, n.$$

Доказательство теоремы проводится аналогично, как и в одномерном случае.

Пример. Пусть

$$x_1^2 + x_2 y^2 = 1.$$

Тогда

$$2x_1 + x_2 y \frac{\partial y}{\partial x_1} = 0,$$

$$y^2 + x_2 2y \frac{\partial y}{\partial x_2} = 0.$$

Отсюда находим $\partial y/\partial x_1$ и $\partial y/\partial x_2$. Аналогично можно найти и частные производные большего порядка.