Задача 11. Методы условной многомерной оптимизации

Спонсор задания — В.И.Гориховский

04 апреля 2025

Ссылки:
https://studfile.net/pnipu/732/folder:1692/#481240

В. Г. Жадан. Методы оптимизации. Часть 2: численные алгоритмы.

Ф. П. Васильев. Методы оптимизации. Том I.

Что нужно сделать:

Реализовать и сравнить методы по метрикам:

- скорость сходимости по числу вычислений оптимизируемой функции
- скорость сходимости по времени
- вероятность нахождения глобального оптимума

Некоторая классификация

- методы возможных направлений (метод проектирования градиента и т.п.)
- методы штрафных функций
- методы барьерных функций
- разное

Есть функция $f : \mathbb{R}^n \to \mathbb{R}$.
Нужно найти точку минимума при ограничениях $x \in D$.

Мы ищем: локальный минимум.

Методы возможных направлений

Идея: определяем *возможное* направление.
Определение. Направление d для точки $x^k \in D$ называется возможным, если существует $\lambda \neq 0$, при котором

$$x^k + \lambda \cdot d \in D, \qquad f(x^k + \lambda \cdot d) < f(x^k).$$

Т.е. есть направление, в котором лучше пойти.

Метод проектирования градиента

Используется проекция градиента на «поверхность ограничений».

Пусть ограничения заданы в виде равенств:
$\psi_j(\mathbf{x}) = 0,\ j = 1 : m;\ \ m < n.$
Возможное направление ℓ — на котором скорость изменения целевой функции максимальна:

$$\frac{df}{d\ell} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{d\ell} \to \max_i .$$

В области D функции ψ_j постоянны, поэтому направление ℓ должно удовлетворять ограничениям:

$$0 = \frac{d\psi_j}{d\ell} = \sum_{i=1}^{n} \frac{\partial \psi_j}{\partial x_i} \frac{dx_i}{d\ell},\ j = 1 : m.$$

Направление и координаты связаны так:
$d\ell^2 = \sum_{i=1}^{n} (dx_i)^2 \Leftrightarrow 1 - \sum_{i=1}^{n} \left(\frac{dx_i}{d\ell}\right)^2 = 0.$

Т.е. имеется задача оптимизации (максимизации) с ограничениями типа равенств и для дифференцируемых функций. Используем метод Лагранжа.

Функция Лагранжа:

$$F\left(\frac{d\mathbf{x}}{d\ell}, \Lambda\right) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{d\ell} + \lambda_0 \left(1 - \sum_{i=1}^{n}\left(\frac{dx_i}{d\ell}\right)^2\right) + \sum_{j=1}^{m}\lambda_j \sum_{i=1}^{n}\frac{\partial \psi_j}{\partial x_i}\frac{dx_i}{d\ell}.$$

Неизвестные: векторы $\frac{d\mathbf{x}}{d\ell}$ и Λ. Берем производные, приравниваем к нулю, выражаем, подставляем, получаем (знаменатель — длина проекции градиента, сравнивается с ε):

$$\frac{dx_i}{d\ell} = \frac{\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m}\lambda_j \frac{\partial \psi_j}{\partial x_i}}{\sqrt{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i} + \sum_{j=1}^{m}\lambda_j\frac{\partial \psi_j}{\partial x_i}\right)^2}},$$

где λ_j ищутся из линейной системы

$$\sum_{i=1}^{n}\frac{\partial \psi_k}{\partial x_i}\sum_{j=1}^{m}\lambda_j\frac{\partial \psi_j}{\partial x_i} = -\sum_{i=1}^{n}\frac{\partial \psi_k}{\partial x_i}\frac{\partial f}{\partial x_i},\ \ k = 1:m.$$

Новая точка вычисляется по формуле:

$$x_i^{k+1} = x_i^k + h^k \cdot \frac{dx_i}{d\ell}, \;\; i = 1 : n.$$

Итого:

- Стартуем с неких $\mathbf{x}^0$ и h^0.
- Если значение целевой функции улучшилось, то считаем дальше, иначе — уменьшаем шаг и повторяем вычисление для данной точки.
- Считаем до тех пор, пока длина проекции градиента не станет меньше ε.

Замечание. Если ограничения типа неравенств: внутри области (когда неравенства строгие) считаем по градиентным методам; на границе (когда часть неравенств становится равенствами) — по методу проектирования градиента.

Метод штрафных функций

Идея: сведение к задаче на безусловный экстремум с помощью преобразования ограничений.

Пусть есть задача:

$$f(\mathbf{x}) \to \min$$
$$\varphi_i(\mathbf{x}) \leqslant 0, \; i = 1 : m_1;$$
$$\psi_i(\mathbf{x}) = 0, \; i = 1 : m_2.$$

Строим вспомогательную функцию $\Theta(\mathbf{x}) = f(\mathbf{x}) + \alpha^k H(\mathbf{x})$, где $H(\mathbf{x})$ — функция штрафа, α^k — параметр штрафа.

Что хочется от $\alpha^k H(\mathbf{x})$

Функция определена и непрерывна на всем $\mathbb{R}^n$.

А также:

- $\alpha^k H(\mathbf{x}) = 0$, $\mathbf{x} \in D$
- $\alpha^k H(\mathbf{x}) > 0$, $\mathbf{x} \notin D$
- $\alpha^{k+1} H(\mathbf{x}) > \alpha^k H(\mathbf{x})$, $\mathbf{x} \notin D$
- $\lim_{k\to\infty} \alpha^k H(\mathbf{x}) = +\infty$, $\mathbf{x} \notin D$

В нашей задаче $H(\mathbf{x}) = H_\varphi(\mathbf{x}) + H_\psi(\mathbf{x})$, при этом:

$$H_\varphi = \begin{cases} 0, & \forall \varphi_i \leqslant 0, \\ > 0, & \exists \varphi_i > 0; \end{cases} \qquad H_\psi = \begin{cases} 0, & \forall \psi_i = 0, \\ > 0, & \exists \psi_i \neq 0. \end{cases}$$

Строить штрафные функции можно по-разному. Например,

$$H_\varphi(\mathbf{x}) = \sum_{i=1}^{m_1} [\max\{0,\ \varphi_i(\mathbf{x})\}]^p\,, \quad H_\psi(\mathbf{x}) = \sum_{i=1}^{m_2} |\psi_i(\mathbf{x})|^{'}\,.$$

где p — какое-нибудь натуральное число. Например, $p = 2$.
От величины α зависит влияние функции штрафа. Но при очень больших α сложно решать задачу минимизации. Поэтому решаем последовательность минимизаций Θ с увеличивающимся значением α.

- Стартуем с точки $\mathbf{x}^0$ (из недопустимой области!), начального α^0. Точность ε.
- Находим $\mathbf{x}_*^k$ — минимум очередной $\Theta(\mathbf{x})$.
- Если $\alpha^k \cdot H(\mathbf{x}_*^k) < \varepsilon$ — конец.
- Иначе увеличиваем α^k и очередную задачу минимизации решаем, начиная с $\mathbf{x}_*^k$.

Метод барьерных функций

Применим к задачам с ограничениями в виде неравенств.
Идея: чтобы последовательно находимые точки не выходили из допустимой области, задача модифицируется так, что при приближении к границе области растет барьер, препятствующий выходу.

Пусть есть задача:

$$f(\mathbf{x}) \to \min$$
$$\varphi_i(\mathbf{x}) \leqslant 0, \ \ i = 1 : m.$$
$$\exists \mathbf{x} : \varphi_i(\mathbf{x}) < 0 \ \forall i$$

Строим вспомогательную функцию $\Theta(\mathbf{x}) = f(\mathbf{x}) + \mu^k B(\mathbf{x})$, где $B(\mathbf{x})$ — функция барьера, μ^k — параметр барьера.

$B(\mathbf{x})$ строится примерно так:

$$B(\cdot) = \begin{cases} \leqslant 0, \ \forall \varphi_i < 0, \\ \infty, \ \exists \varphi_i = 0. \end{cases}$$

Строить барьерные функции можно по-разному. Например,

$$B(\mathbf{x}) = \sum_{i=1}^{m} \frac{-1}{\varphi_i(\mathbf{x})}.$$

От величины μ зависит влияние функции барьера. Но при очень маленьких μ сложно решать задачу минимизации. Поэтому решаем последовательность минимизаций Θ с уменьшающимися значениями μ.

- Стартуем с точки $\mathbf{x}^0$ (из строго допустимой области), начального μ^0. Точность ε.
- Находим $\mathbf{x}_*^k$ — минимум очередной $\Theta(\mathbf{x})$. Проверяем, что не вышли из допустимой области!
- Если $\mu^k \cdot B(\mathbf{x}_*^k) < \varepsilon$ — конец.
- Иначе уменьшаем μ^k и очередную задачу минимизации решаем, начиная с $\mathbf{x}_*^k$.

Метод модифицированных функций Лагранжа

Пусть есть задача:

$$f(\mathbf{x}) \to \min \qquad \varphi_i(\mathbf{x}) = 0, \ i = 1 : m.$$

Строим не функцию Лагранжа ($L(\mathbf{x}, u) = f(\mathbf{x}) + (\varphi(\mathbf{x}), u), \ u \in \mathbb{R}^m$), а модифицированную функцию Лагранжа:

$$M(\mathbf{x}, u, t) = f(\mathbf{x}) + \sum_{i=1}^{m} \left[u_i \varphi_i(\mathbf{x}) + \frac{t}{2} (\varphi_i(\mathbf{x}))^2 \right].$$

Т.е. добавили штраф к функции Лагранжа.
Стационарные точки функции Лагранжа и модифицированной функции Лагранжа совпадают.

Итерационный процесс:

- Пусть задано начальное $u^0 \in \mathbb{R}^m$ и выбрано и зафиксировано достаточно большое t.
- $\mathbf{x}^k = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^n} M(\mathbf{x},\, u^k,\, t)$,
  $u^{k+1} = u^k + t\varphi(\mathbf{x}^k)$
- Если $|\mathbf{x}^k - \mathbf{x}^{k-1}| < \varepsilon$ — конец.

Задание 11. Методы условной многомерной оптимизации

Список методов (выбрать три):

- метод проектирования градиента
- метод штрафных функций
- метод барьерных функций
- метод модифицированных функций Лагранжа