Численная оптимизация

Функции многих переменных

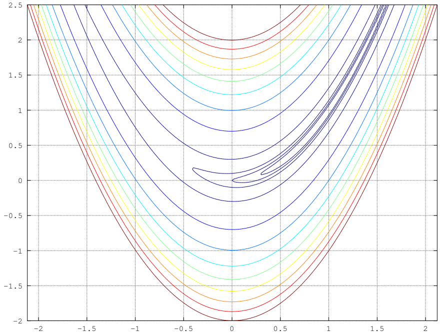
12 ноября 2012 г.

- Методы прямого поиска:
  - поиск по симплексу;
  - метод Хука–Дживса;
  - метод сопряжённых направлений Пауэлла.
- Градиентные методы:
  - метод Коши;
  - метод Ньютона;
  - метод Марквардта;
  - квазиньютоновские методы;
- Методы решения задач о наименьших квадратах.

Тестовые функции

Функция Розенброка

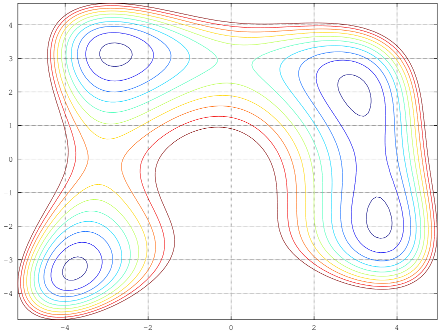
$$f(x) = \sum_{i=1}^{n-1} \left((1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2\right)$$



Тестовые функции

Функция Химмельблау

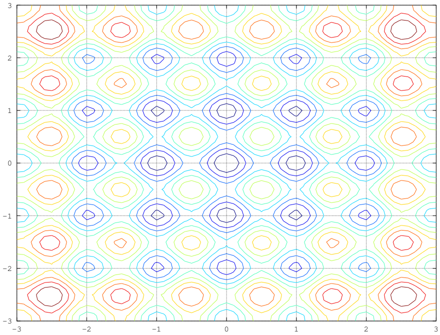
$$f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$



Тестовые функции

Функция Растригина

$$f(x) = An + \sum_{i=1}^{n-1} \left(x_i^2 - A\cos(2\pi x_i)\right)$$



Поиск по симплексу (S^2-метод)

Регулярный симплекс в N-мерном пространстве

Многогранник, образованный $N+1$ равноотстоящими друг от друга точками–вершинами.

$x^{(0)}$ — базовая (начальная) точка

$$x^{(i)} = \begin{cases} x_j^{(0)} + \delta_1, & j \neq i, \\ x_j^{(0)} + \delta_2, & j = i, \end{cases} \quad i, j = 1, 2, \ldots, N.$$

$$\delta_1 = \frac{\sqrt{N+1} + N - 1}{N\sqrt{2}}\alpha, \quad \delta_2 = \frac{\sqrt{N+1} - 1}{N\sqrt{2}}\alpha.$$

α — масштабный коэффициент

Поиск по симплексу (S^2-метод)

$x^{(j)}$ — отражаемая точка

«Центр тяжести» оставшихся точек:

$$x_c = \frac{1}{N} \sum_{\substack{i=0,1,\ldots,N \\ i \neq j}} x^{(i)}$$

$$x = x^{(j)} + \lambda(x_c - x^{(j)})$$

Новая точка:

$$x^{(j)}_{\text{нов}} = 2x_c - x^{(j)}_{\text{пред}}$$

Поиск по симплексу (S^2-метод)

1. "Накрытие" точки минимума.
2. Циклическое движение

$$M = 1.65N + 0.05N^2$$

3. Критерий окончания поиска

Поиск по симплексу (S^2-метод)

Преимущества

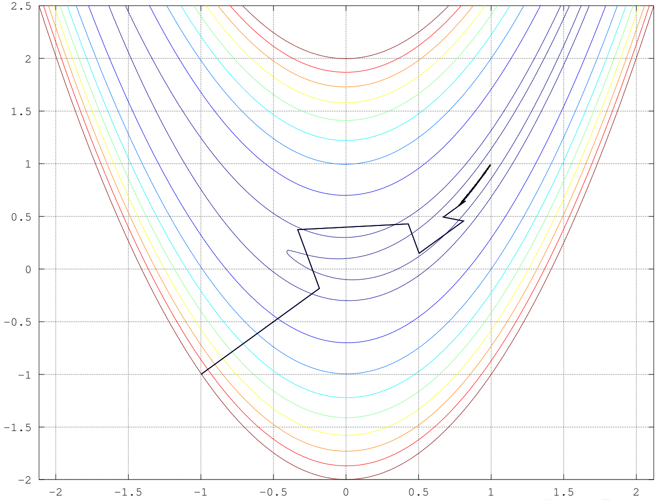
1. Сравнительная простота расчётов и логической структуры метода.
2. Невысокий уровень требований к объёму машинной памяти.
3. Небольшое число параметров: масштабный множитель α, коэффициент уменьшения значения α и параметры окончания поиска.
4. Алгоритм эффективен даже в случае большой ошибки вычисления целевой функции.

Поиск по симплексу (S^2-метод)

Недостатки

1. Возможны проблемы с масштабированием.
2. Медленная работа алгоритма.
3. Не существует простого способа расширения симплекса, не требующего пересчёта значений целевой функции во всех точках образца.

Количество итераций: 8277.



Поиск по деформируемому многограннику

$$x = x^{(j)} + (1+\theta)\cdot(x_c - x^{(j)})$$

$$(1, 0.5, 2)$$

$$(1, 0.25, 2.5)$$

Количество итераций при минимизации функции Розенброка: 398 (1211)

Метод поиска Хука–Дживса

- Исследующий поиск
- Поиск по образцу

Метод поиска Хука–Дживса

1. Определить: начальную точку $x^{(0)}$; приращения Δ_i, $i = 1, 2, \ldots, N$; коэффициент уменьшения шага $\alpha > 1$; параметр окончания поиска $\varepsilon > 0$;
2. Провести исследующий поиск.
3. Если исследующий поиск был удачный, то перейти к шагу 5.
4. Если $\|\Delta x\| < \varepsilon$, то прекратить поиск.
   Иначе — уменьшить приращения по формуле

   $$\Delta_i \leftarrow \Delta_i / \alpha, \quad i = 1, \ldots, N.$$

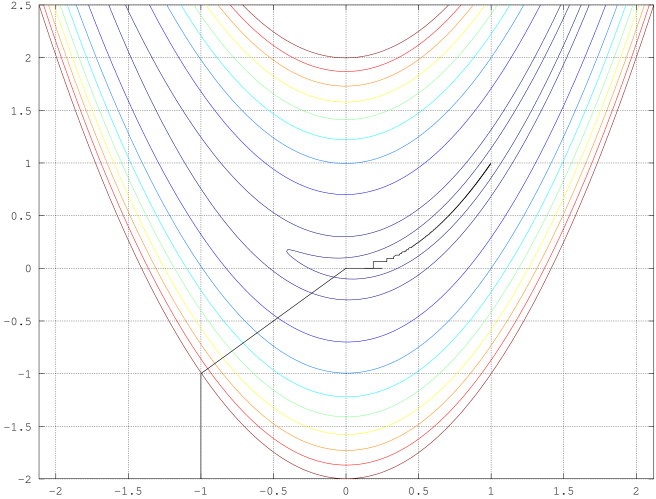
   Перейти к шагу 2.

Метод поиска Хука–Дживса

5. Провести поиск по образцу
$$x_p^{(k+1)} = x^{(k)} + (x^{(k)} - x^{(k-1)}).$$
6. Провести исследующий поиск, используя $x_p^{(k+1)}$ в качестве базовой точки; пусть $x^{(k+1)}$ — полученная в результате точка.
7. Если $f(x^{(k+1)}) < f(x^{(k)})$, то $x^{(k-1)} \leftarrow x^{(k)}$, $x^{(k)} \leftarrow x^{(k+1)}$ и перейти к шагу 5.
   Иначе перейти к шагу 4.

Метод поиска Хука–Дживса

Минимизация функции Розенброка: 1609 итераций.



Метод сопряжённых направлений Пауэлла

Метод ориентирован на решение задач с квадратичными целевыми функциями.

1. Квадратичная функция представляет собой простейший вид нелинейных функций, для которых может быть построена задача безусловной оптимизации.
2. В окрестности точки оптимума любую гладкую нелинейную функцию можно аппроксимировать квадратичной функцией.

Метод сопряжённых направлений Пауэлла

Если квадратичная функция N переменных приведена к виду суммы полных квадратов, то её оптимум может быть найден в результате реализации N одномерных поисков по преобразованным координатным направлениям.

Метод сопряжённых направлений Пауэлла

C — симметричная $N \times N$-матрица.

Направления $s^{(1)}, s^{(2)}, \dots, s^{(r)}$, $r \leqslant N$, называются C-сопряжёнными, если эти направления линейно независимы и

$$s^{(i)\top} C x^{(j)} = 0 \text{ для всех } i \neq j.$$

C-сопряжённые направления задаются системой векторов T, которая приводит C к диагональному виду:

$$T^\top C T = D$$

Метод сопряжённых направлений Пауэлла

Свойство параллельного пространства

Пусть заданы квадратичная функция $q(x)$, две произвольные несовпадающие точки $x^{(1)}$ и $x^{(2)}$, а также направление d. Если точка $y^{(1)}$ минимизирует $q(x^{(1)} + \lambda d)$, а точка $y^{(2)}$ минимизирует $q(x^{(2)} + \lambda d)$, то направление $y^{(2)} - y^{(1)}$ сопряжено с d.

$$q(x) = a + b^\top x + \frac{1}{2} x^\top C x$$

$$x = x^{(1)} + \lambda d, \qquad \frac{\partial q}{\partial \lambda} = \frac{\partial q}{\partial x}\frac{\partial x}{\partial \lambda} = (b^\top + x^\top C)d$$

$$({y^{(1)}}^\top C + b^\top)d = 0, \quad ({y^{(2)}}^\top C + b^\top)d = 0$$

$$(y^{(2)} - y^{(1)})^\top C d = 0$$

Метод сопряжённых направлений Пауэлла

Обобщённое свойство параллельного пространства

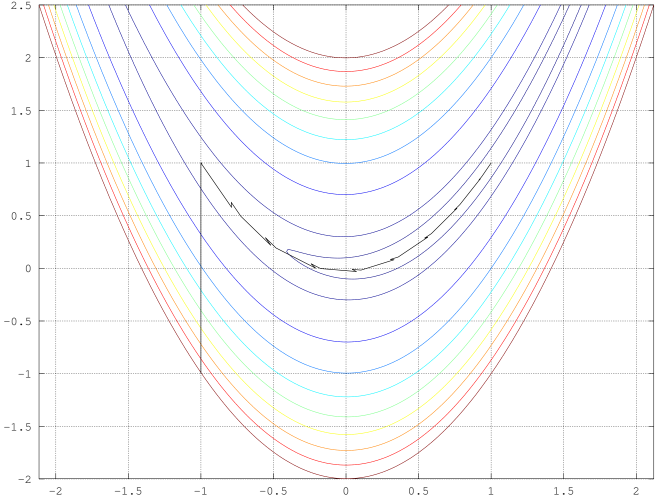
Если точка $y^{(1)}$ найдена в результате поиска из точки $x^{(1)}$ вдоль каждого из M $(< N)$ сопряжённых направлений, а точка $y^{(2)}$ получена в результате поиска из точки $x^{(2)}$ вдоль каждого из тех же M сопряжённых направлений $s^{(1)}, s^{(2)}, \ldots, s^{(M)}$, то вектор $y^{(2)} - y^{(1)}$ задаёт направление, сопряжённое *со всеми* выбранными M направлениями.

Метод сопряжённых направлений Пауэлла

1. Задать начальную точку $x^{(0)}$ и систему N линейно независимых направлений.
2. Минимизировать $f(x)$ при последовательном движении по $N+1$ направлениям. При этом полученая ранее точка минимума берётся в качестве исходной, а направление $X^{(N)}$ используется как при первом, так и при последнем поиске.
3. Определить новое сопряжённое направление с помощью обобщённого свойства параллельного подпространства.
4. Заменить $s^{(1)}$ на $s^{(2)}$ и т.д. Заменить $x^{(N)}$ сопряжённым направлением и перейти к шагу 2.

Метод сопряжённых направлений Пауэлла

Минимизация функции Розенброка: 10 итераций.



Метод Коши (метод наискорейшего спуска)

$$f(x) = f(\tilde{x}) + \nabla f(\tilde{x}) \Delta x + \cdots$$

$$s(\tilde{x}) = -\nabla f(\tilde{x})$$

Простейший градиентный метод

$$x^{(k+1)} = x^{(k)} - \alpha \cdot \nabla f(x^{(k)})$$

Недостатки:

1. Необходимо выбрать подходящее значение α.
2. Низкая скорость сходимости в окрестности точки минимума.

Метод Коши

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \cdot \nabla f(x^{(k)})$$

Метод Ньютона

$$f(x) = f(\tilde{x}) + \nabla f(\tilde{x})\Delta x + \frac{1}{2}\Delta x^\top \nabla^2 f(\tilde{x})\Delta x + \cdots$$

$$\hat{f}(x, \tilde{x}) = f(\tilde{x}) + \nabla f(\tilde{x})\Delta x + \frac{1}{2}\Delta x^\top \nabla^2 f(\tilde{x})\Delta x$$

$$\nabla \hat{f}(x, \tilde{x}) = \nabla f(\tilde{x}) + \nabla^2 f(\tilde{x})\Delta x$$

$$s(\tilde{x}) = -\nabla^2 f(\tilde{x}) \nabla f(\tilde{x})$$

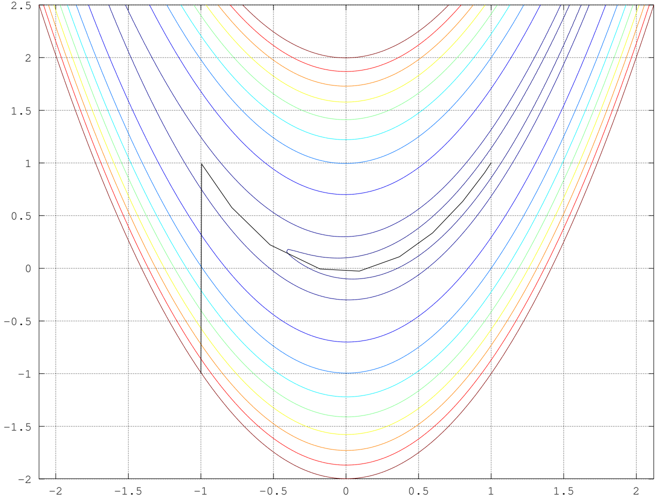
Метод Ньютона

$$x^{(k+1)} = x^{(k)} - \nabla^2 f(x^k)^{-1} \cdot \nabla f(x^{(k)})$$

Модифицированный метод Ньютона

$$x^{(k+1)} = x^{(k)} - \alpha^{(k)} \cdot \nabla^2 f(x^k)^{-1} \cdot \nabla f(x^{(k)})$$

Минимизация функции Розенброка: 13 итераций.



Метод Марквардта

Объединение методов Коши и Ньютона.

Направление поиска

$$s(x^{(k)}) = -\left(\nabla^2 f(x^{(k)}) + \lambda^{(k)} E\right)^{-1} \cdot \nabla f(x^{(k)})$$

Метод Марквардта

1. Задать: $x^{(0)}$ — начальное приближение, M — максимальное число итераций, ε — параметр сходимости.
2. $k = 0$, $\lambda^{(0)} = 10^4$
3. Вычислить коэффициенты $\nabla f(x^{(k)})$.
4. Если $\|\nabla f(x^{(k)})\| < \varepsilon$, то перейти к шагу 11.
5. Если $k \geqslant M$, то перейти к шагу 11.
6. Вычислить
$$s(x^{(k)}) = -\left(\nabla^2 f(x^{(k)}) + \lambda^{(k)} E\right)^{-1} \cdot \nabla f(x^{(k)}).$$
7. $x^{(k+1)} = x^{(k)} + s(x^{(k)})$

8. Если $f(x^{(k+1)}) \geqslant f(x^{(k)})$, то перейти к шагу 10.
9. $\lambda^{(k+1)} = \lambda^{(k)}/2$, $k \leftarrow k+1$.
   Перейти к шагу 3.
10. $\lambda^{(k)} \leftarrow 2\lambda^{(k)}$.
    Перейти к шагу 6.
11. Вывод результатов. Закончить вычисления.

Квазиньютоновские методы

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} \cdot s(x^{(k)})$$

$$A^{(k)} s(x^{(k)}) = -\nabla f(x^{(k)})$$

Аппроксимация матрицы Гессе

$$A^{(k+1)} = A^{(k)} + A_c^{(k)}$$

$A_c^{(k)}$ — корректирующая матрица

$$s_k = x^{(k+1)} - x^{(k)}, \quad y_k = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$$

Квазиньютоновское условие

$A^{(k+1)}$ правильно отражает кривизну f вдоль s_k:

$$A^{(k+1)} s_k = y_k$$

Поправка ранга один

$$A^{(k+1)} = A^{(k)} + uv^\top$$

$$A^{(k+1)}s_k = (A^k + uv^\top)s_k = y_k$$

$$u(v^\top s_k) = y_k - A^{(k)}s_k$$

Вектор u должен быть параллелен $y_k - A^{(k)}s_k$

Если задан вектор v (не ортогональный s_k), то

$$A^{(k+1)} = A^{(k)} + \frac{1}{v^\top s_k}(y_k - A^{(k)}s_k)v^\top$$

Для сохранения симметрии матрицы вектор v должен быть коллинеарен u.

Поправка ранга 1

$$A^{(k+1)} = A^{(k)} + \frac{1}{(y_k - A^{(k)} s_k)^\top s_k}(y_k - A^{(k)} s_k)(y_k - A^{(k)} s_k)^\top$$

$y_k - A^{(k)} s_k$ и $(y_k - A^{(k)} s_k)^\top s_k$ предполагаются ненулевыми

$$A_1 = A^{(k)} + \frac{1}{v^\top s_k}(y_k - A^{(k)} s_k) v^\top$$

$$A_2 = \frac{1}{2}(A_1 + A_1^\top)$$

$$A_{2j+1} = A_{2j} + \frac{1}{v^\top s_k}(y_k - A_{2j} s_k) v^\top$$

$$A_{2j+2} = \frac{1}{2}(A_{2k+1} + A_{2j+1}^\top)$$

$$j = 0, 1, \ldots$$

Поправка ранга 2

Для любого вектора v, не ортогонального s_k:

$$A^{(k+1)} = A^{(k)} + \\ + \frac{1}{v^\top s_k}\left((y_k - A^{(k)} s_k)v^\top + v(y_k - A^{(k)} s_k)^\top\right) - \\ - \frac{(y_k - A^{(k)} s_k)^\top s_k}{(v^\top s_k)^2} v v^\top$$

- $v = y_k - A^{(k)} s_k$ — формула с поправкой ранга 1;
- $v = s_k$ — симметричная формула Пауэлла–Бройдена;
- $v = y_k$ — формула Дэвидона–Флетчера–Пауэлла.

Формула Дэвидона–Флетчера–Пауэлла:

$$A^{(k+1)} = A^{(k)} - \frac{1}{s_k^\top A^{(k)} s_k} A^{(k)} s_k s_k^\top A^{(k)} + \\ + \frac{1}{y_k^\top s_k} y_k y_k^\top + (s_k^\top A^{(k)} s_k) w_k w_k^\top,$$

$$w_k = \frac{1}{y_k^\top s_k} y_k - \frac{1}{s_k^\top A^{(k)} s_k} A^{(k)} s_k$$

$$w_k \perp s_k$$

Семейство однопараметрических формул:

$$A^{(k+1)} = A^{(k)} - \frac{1}{s_k^\top A^{(k)} s_k} A^{(k)} s_k s_k^\top A^{(k)} + \\ + \frac{1}{y_k^\top s_k} y_k y_k^\top + \phi_k (s_k^\top A^{(k)} s_k) w_k w_k^\top,$$

$$w_k = \frac{1}{y_k^\top s_k} y_k - \frac{1}{s_k^\top A^{(k)} s_k} A^{(k)} s_k$$

Формула Бройдена–Флетчера–Гольдфарба–Шанно (BFGS)

$$A^{(k+1)} = A^{(k)} - \frac{1}{s_k^\top A^{(k)} s_k} A^{(k)} s_k s_k^\top A^{(k)} + \frac{1}{y_k^\top s_k} y_k y_k^\top$$

$$A^{(k)} > 0 \quad \Rightarrow \quad \exists R \colon A^{(k)} = R^\top R$$

Для формулы BFGS:

$$A^{(k+1)} = R^\top W R$$

$$W = I - \frac{1}{\bar{s}^\top \bar{s}} \bar{s}\bar{s}^\top + \frac{1}{\bar{y}^\top \bar{s}} \bar{y}\bar{y}^\top$$

$$\bar{s} = R s_k, \quad \bar{y} = (R^\top)^{-1} y_k$$

$$\lambda_1 + \lambda_1 = \frac{\bar{y}^\top \bar{s} + \bar{y}^\top \bar{y}}{\bar{y}^\top \bar{s}}, \quad \lambda_1 \lambda_2 = \frac{\bar{y}^\top \bar{s}}{\bar{s}^\top \bar{s}}$$

Задача о наименьших квадратах

$$F(x) = \frac{1}{2}\sum_{i=1}^{m} f_i(x)^2 = \|f(x)\|_2^2 \to \min$$

$f(x)$ — нелинейная векторная функция с m компонентами
$\|f(x)\|_2$ — невязка в точке x

Задача о наименьших квадратах

$$\min_{x\in\mathbb{R}^n} \int_{t_0}^{t_1} \big(\phi(x,t) - \mathcal{F}(t)\big)^2 \, dt$$

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} \big(\bar{\phi}(x,t_i) - \bar{\mathcal{F}}(t_i)\big)^2$$

Задача о наименьших квадратах

$J(x)$ — $m \times n$-матрица Якоби для $f(x)$
$G_i(x)$ — матрица Гессе для $f_i(x)$

Градиент функции $F(x)$

$$g(x) = J(x)^\top f(x)$$

Матрица Гессе функции $F(x)$

$$G(x) = J(x)^\top J(x) + Q(x), \quad Q(x) = \sum_{i=1}^{m} f_i(x) G_i(x)$$

Метод Ньютона:

$$(J_k^\top J_k + Q_k)p_k = -J_k^\top f_k$$

Если при приближении x_k к оптимуму $\|f_k\|$ приближается к нулю, то матрица Q_k также будет приближаться к нулевой.

Аппроксимация ньютоновского направления

$$J_k^\top J_k p_k = -J_k^\top f_k$$

Линейная задача о наименьших квадратах

$$\min_{p \in \mathbb{R}^n} \frac{1}{2}\|J_k p + f_k\|_2^2$$

$$(J_k^\top J_k + \lambda_k I)p_k = -J_k^\top f_k$$

λ_k — некоторое неотрицательное число

$x_{k+1} = x_k + p_k$

$$\min_{p \in \mathbb{R}^n} \frac{1}{2}\|J_k p + f_k\|_2^2$$

$$\|p\|_2 \leqslant \Delta$$

Δ — параметр, связанный с λ_k

$$(J_k^\top J_k + M_k)p_k = -J_k^\top f_k$$

Квазиньютоновское условие

$$(J_{k+1}^\top J_{k+1} + M_{k+1})s_k = y_k$$

$s_k = x_{k+1} - x_k$
$y_k = J_{k+1}^\top f_{k+1} - J_k^\top f_k$

На основе BFGS:

$$M_{k+1} = M_k - \frac{1}{s_k^\top W_k s_k} W_k s_k s_k^\top W^\top + \frac{1}{y_k^\top s_k} y_k y_k^\top$$

$$W_k = J_{k+1}^\top J_{k+1} + M_k$$