Лекция 15

Ilya Yaroshevskiy

21 ноября 2024 г.

Содержание

1 Метод Ньютона

Если целевая функция дважды дифференцируема в пространстве E_n в процессе поиска можно использовать не только информации о градиенте, но и матрице Гесса. Для оптимизируемой функции $f(x) \in E_n$ разложение по формуле Тейлора:

$$f(x) = f(x^{k-1}) + (\nabla f(x^{k-1}), x - x^{k-1}) + \frac{1}{2}(H(x^{k-1})(x - x^{k-1}), x - x^{k-1}) + o(|x - x^{k-1}|)$$

$$\varphi_k(x) = f(x^{k-1}) + (\nabla f(x^{k-1}), x - x^{k-1}) + \frac{1}{2}(H(x^{k-1})(x - x^{k-1}), x - x^{k-1})$$

$$\nabla \varphi_k(x) = \nabla f(x^{k-1}) + H(x^{k-1})(x - x^{k-1})$$

$$\left[\ \nabla(a^T x) = a; \quad \nabla(x^T A x) = 2Ax\ \right]$$

$$x^k = x^{k-1} - H^{-1}(x^{k-1})\nabla f(x^{k-1}) \quad k \in \mathbb{N} \tag{1}$$

$\tilde{x}^k$ — вспомогательная точка для построения релаксационной последовательности $\{x^k\}$

$$x^k = x^{k-1} + \alpha_k(\tilde{x}^k - x^{k-1}) = x^{k-1} + \alpha_k p^k \quad \alpha_k > 0 \tag{2}$$

$p^k = \tilde{x}^k - x^{k-1}$ — направление спуска

$$p^k = -H^{-1}(x^{k-1})\nabla f(x^{k-1}) \tag{3}$$

$$(\nabla f(x^{k-1}), p^k) = -(\nabla f(x^{k-1}), H^{-1}(x^{k-1})\nabla f(x^{k-1})) < 0$$

H — положительно определена $\implies$ H^{-1} — положительно определена Выбор $\alpha_k \to$

1. $\varphi_k(\alpha) = f(x^{k-1} + \alpha p^k) \to \min$
2. исчерпывающий спуск по p^k
3. дробление α_k

$f(x)$ — квадратичная функция с положительно определенной матрицей A. x^0 — p^k — ньютоновское направление $\implies$ точка минимума за одну итерацию

$$f(x) = \frac{1}{2}(Ax, x) + (b, x); \quad A = H$$

$$p^1 = -A^{-1}\nabla f(x^0) = -A^{-1}(Ax^0 + b) = -x^0 - A^{-1}b$$

$$x^1 = x^0 + p^1 = -A^{-1}b$$

1.1 Сходимость метода Ньютона

- Зависит от x^0. Если $f(x)$ сильно выпуклая и $\forall x, y \in E_n$

  $$H(x): \ \|H(x) - H(y)\| \leq L|x - y| \quad L > 0$$

  и удачный выбор x^0 (x^0 близок к x^*), то метод Ньютона при $\alpha_k = 1$ в 2 обладает квадратичной сходимости

  $$|x^k - x^*| \leq C|x^{k-1} - x^*|^2 \quad C = const$$

  Если $f(x)$ не является сильно выпуклой или начальное приближение x^0 далеко от x^* $\implies$ метод Ньютона может расходиться

Алгоритм 1 Метод Ньютона

повторять
  Вычислить $\nabla f(x)$
  $H = \nabla^2 f(x)$
  Решить СЛАУ: $Hp^k = -\nabla f(x)$
  $x^k = x^{k-1} + p^k$
до тех пор пока $\|x^k - x^{k-1}\| < \varepsilon$ ($\|p^k\| < \varepsilon$)
$x^* = x^k$

1.2 Метод Ньютона с одномерным поиском

$x^k \to$ одномерный поиск в направлении p^k

$$\alpha_k = \min_{\alpha} f(x^k + \alpha p^k) \tag{4}$$

$$H(x^{k-1})p^k = -\nabla f(x^{k-1})$$

Пока $\|x^k - x^{k-1}\| \geq \varepsilon$ — итерации продолжать

Алгоритм 2 Метод Ньютона с одномерным поиском

повторять
  Вычислить $\nabla f(x)$
  $H = \nabla^2 f(x)$
  Решить СЛАУ: $Hp^k = -\nabla f(x)$
  $\alpha_k = \min_{\alpha} f(x^{k-1} + \alpha p^k)$
  $x^k = x^{k-1} + \alpha_k p^k$
до тех пор пока $\|x^k - x^{k-1}\| < \varepsilon$ ($\|p^k\| < \varepsilon$)
$x^* = x^k$

Метод надежнее обычного метода Ньютона, но его эффективность существенно зависит от того, является ли p^k направлением спуска

1.3 Метод Ньютона с направлением спуска

Если p^k — направление спуска: $(p^k)^T \nabla f(x^k) < 0$. Если $(p^k)^T \nabla f(x^k) > 0$, то p^k — не является направлением спуска, в этом случае использовать антиградиент $-\nabla f(x^k)$.

$$H(x^k)p^k = -\nabla f(x^k) \implies p^k = \begin{cases} p^k & (p^k)^T \nabla f(x^k) < 0 \\ -\nabla f(x^k) & (p^k)^T \nabla f(x^k) > 0 \end{cases} \tag{5}$$

$$x^1 = x^0 + \alpha_0 p^0 \quad p^0 = -\nabla f(x^0) \quad \alpha_0 = \min_{\alpha} f(x^0 + \alpha p^0) \tag{6}$$

$(k > 1)$

$$x^k = x^{k-1} + \alpha_k p^k \quad \alpha_k = \min_{\alpha} f(x^{k-1} + \alpha p^{k-1}) \tag{7}$$

$$\|x^k - x^{k-1}\| > \varepsilon$$

Алгоритм 3 Метод Ньютона с направлением спуска

повторять
 Вычислить $\nabla f(x)$
 $H = \nabla^2 f(x)$
 Решить СЛАУ: $Hp^k = -\nabla f(x)$
 если $(p^k)^T \nabla f(x^k) > 0$ **тогда**
  $p^k = -\nabla f(x^k)$
 конец если
 $\alpha_k = \min_\alpha f(x^{k-1} + \alpha p^k)$
 $x^k = x^{k-1} + \alpha_k p^k$
до тех пор пока $\|x^k - x^{k-1}\| < \varepsilon$ $(\|p^k\| < \varepsilon)$
$x^* = x^k$

1.4 Метод Ньютона с дроблением шага

$$f(x^{k-1}) - f(x^k) \geq -\omega \alpha_k (\nabla f(x^{k-1}), p^k) \quad \omega \in \left(0, \frac{1}{2}\right) \tag{8}$$

Начальное $\alpha_k = 1 \implies$ проверим условие 8, если нарушено, то α_k — корректировка, $\alpha_k^* \nu$, ν — коэффициент,снова проверка $\ldots$, $\nu \in (0,1)$

Алгоритм 4 Метод Ньютона с дроблением шага

Ввод $\nu \in (0,1)$
повторять
 Вычислить $\nabla f(x)$
 $H = \nabla^2 f(x)$
 Решить СЛАУ: $Hp^k = -\nabla f(x)$
 если $(p^k)^T \nabla f(x^k) > 0$ **тогда**
  $p^k = -\nabla f(x^k)$
 конец если
 $\alpha_k = \min_\alpha f(x^{k-1} + \alpha p^k)$
 пока $f(x^{k-1}) - f(x^k) < -\omega \alpha_k (\nabla f(x^{k-1}), p^k)$ **делать**
  $\alpha_k = \alpha_k \cdot \nu$
 конец пока
 $x^k = x^{k-1} + \alpha_k p^k$
до тех пор пока $\|x^k - x^{k-1}\| < \varepsilon$ $(\|p^k\| < \varepsilon)$
$x^* = x^k$

1.5 О сходимости

Квадратичная сходимость: $|x^k - x^*| \leq L|x^{k-1} - x^*|^2$

1. $x^0 \quad H(x^0), H^{-1}$

$$p^k = -H^{-1}(x_0)\nabla f(x^{k-1}) \tag{9}$$

   Если $H(x^0)$ не является положительно определенной, то в 9

$$\tilde{H}^1 = \tau_1 I + H(x^0)$$

   , $\tau_1 \implies \tilde{H}^1$ — положительно определенная. Этот подход дает линейную скорость сходимости $|x_k - x^*| \leq q|x^{k-1} - x^*|$, $q > 0$

2. H — обновляется через фиксированное количество итераций

   $m \;\; H^{-1}(x^0)$

   $2m \;\; H^{-1}(x^m)$

   Этот подход имеет сверхлинейную скорость сходимости: $|x^{km} - x^*| \leq C|x^{(k-1)m} - x^*|^{m+1}$, $C > 0$

2 Метод Марквардта

Комбинация методов наискорейшего спуска и метода Ньютона

1. движение из x^0 в направлении $\nabla f(x)$ приводит к существенному уменьшению $f(x)$
2. направление эффективного поиска в окрестности точки x^* определяется по методу Ньютона

2.1 Идея метода

$$(H(x) + \tau I)\, p^k = -\nabla f(x) \tag{10}$$

, где τ — скалярный параметр, I — единичная матрица

При большом τ в 10 матрицей $H(x)$ можно пренебречь, тогда получим $\tau I p^k = -\nabla f(x)$, то есть $p^k = \frac{-\nabla f(x)}{\tau}$ — совпадает с направлением антиградиента — направлением наискорейшего спуска

При $\tau \to 0$ в 10 можно пренебречь τI, тогда $H(x)p^k = -\nabla f(x)$ — метод Ньютона

При промежуточном τ направление p^k лежит между направлением антиградиента и направлением метода Ньютона

2.2 Реализация

Выбираем $\tau \gg 1$. В процессе поиска $\tau_k = \tau_{k-1} \cdot \beta$, $\beta \in (0, 1)$. В этом случае на начальных шагах выполняются итерации по методу наискорейшего спуска, а на конечных — по методу Ньютона

$$\begin{gathered} x^k = x^{k-1} + p^k \\ \left(H(x^{k-1}) + \tau_{k-1} I\right) p^k = -\nabla f(x^{k-1}) \\ \tau_k = \tau_{k-1} \cdot \beta \end{gathered} \tag{11}$$

$\left\|x^k - x^{k-1}\right\| < \varepsilon$ —условие останова

τ_k — позволяет изменять направление поиска и регулировать длину шага. Если шаг слишком большой и $f(x^{k+1}) > f(x^k)$, то $\tau_k = \frac{\tau_k}{\beta}$ и вновь применяют 11, но уже с меньшим p^k

Метод Марквардта позволяет устранить недостаток метода Ньютона, связанный с возможной плохой обусловленностью СЛАУ:

- для $I \implies \mathrm{cond}(I) = 1$
- для $H \implies \mathrm{cond}(H) \geq 1$

В общем случае H может быть вырожденной $\implies$ СЛАУ 3 не имеет решения, однако СЛАУ 11 лишена этого недостатка

В точках, далеких от минимума, H является, как правило, плохо обусловленной, но в этих точках $\tau \gg 1$, поэтому

$$H + \tau U \approx \tau I$$

$$\mathrm{cond}(H + \tau I) \approx \mathrm{cond}(\tau I) = \|\tau U\| \cdot \left\|(\tau I)^{-1}\right\| = \tau \cdot 1 \cdot \frac{1}{\tau} = 1$$

поэтому в 11 СЛАУ хорошо обусловлена

Алгоритм 5 Метод Марквардта

Ввод $x^0, \tau_0, \beta, \varepsilon$
  $x = x_0$, вычислить $f(x)$
  повторять
    вычислить $\nabla f(x),\ H(x),\ \tau = \tau_0 \cdot \beta$
    повторять
      $\tau = \frac{\tau}{\beta}$
      СЛАУ: $(H(x) + \tau I)p = -\nabla f(x)$
      $y = x + p,\ f(y)$
    до тех пор пока $f(y) > f(x)$
    $x = y,\ f(x) = f(y),\ \tau_0 = \tau_0 \cdot \beta$
  до тех пор пока $\|p\| > \varepsilon$

Метод характеризуется:

- относительной простотой
- свойством убывания $f(x)$ при переходе от итерации к итерации
- высокой скоростью сходимости в окрестности x^*
- отсутствием процедуры одномерного поиска

В некоторых случаях метод Марквардта дополняют одномерным поиском, тогда

$$x^k = x^{k-1} + \alpha_k p^k$$

$$\alpha_k = \min_{\alpha} f(x^{k-1} + \alpha p^k)$$

$$\left(H(x^{k-1}) + \tau_k I\right) p^k = -\nabla f(x^{k-1})$$

$$\tau_k = \tau_{k-1} \cdot \beta$$

$\left\|x^k - x^{k-1}\right\| < \varepsilon$ —условие останова

2.3 Другая вариация метода

$$\left(H(x) + \tau I\right) p^k = -\nabla f(x)$$

Начать с $\tau = 0$
На каждой итерации проверять условие:

- $H(x) + \tau I > 0$ — положительно определена?
- если отрицательно определена, то $\tau = \max(1, 2\tau)$ и снова проверка

Замечание. Условие положительно определенной матрицы $H(x) + \tau I$ проверяется с помощью алгоритма Холецкого:

$$H(x) + \tau I = LL^T$$

, где L — нижнетреугольная матрица. Если разложение $H(x) + \tau I$ на LL^T возможно, то матрица > 0
Сложность этого алгоритма — $\frac{n^3}{3}$, сложность метода — число запусков алгоритма Холецкого

Метод Марквардта широко используется при решении задач, в которых целевая функция записывается в виде суммы квадратов