Лекция 4

МЕТОДЫ ПЕРВОГО ПОРЯДКА

Постановка задачи

Пусть дана функция $f(x)$, ограниченная снизу на множестве R^n и имеющая непрерывные частные производные во всех его точках.

Требуется найти локальный минимум функции $f(x)$ на множестве допустимых решений $X = R^n$, т.е. найти такую точку $x^* \in R^n$, что

$$f(x^*) = \min_{x \in R^n} f(x).$$

А. МЕТОД ГРАДИЕНТНОГО СПУСКА С ПОСТОЯННЫМ ШАГОМ

Стратегия поиска

Стратегия решения задачи состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу

$$x^{k+1} = x^k - t_k \nabla f\left(x^k\right),\ k = 0,1,\ldots ,$$

где точка x^0 задается пользователем; $\nabla f\left(x^k\right)$ – градиент функции $f(x)$, вычисленный в точке x^k; величина шага t_k задается пользователем и остается постоянной до тех пор, пока функция убывает в точках последовательности, что контролируется путем проверки выполнения условия $f\left(x^{k+1}\right) - f\left(x^k\right) < 0$ или $f\left(x^{k+1}\right) - f\left(x^k\right) < -\varepsilon \left\| \nabla f\left(x^k\right)\right\|^2, 0 < \varepsilon < 1$.

Построение последовательности $\left\{x^k\right\}$ заканчивается в точке x^k, для которой $\left\| \nabla f\left(x^k\right)\right\| < \varepsilon_1$, где ε_1 – заданное малое положительное число, или $k \ge M$, где M – предельное число итераций, или при двукратном одновременном выполнении двух неравенств $\left\|x^{k+1} - x^k\right\| < \varepsilon_2$, $\left|f\left(x^{k+1}\right) - f\left(x^k\right)\right| < \varepsilon_2$, где ε_2 – малое положительное число.

Вопрос о том, может ли точка x^k рассматриваться как найденное приближение искомой точки минимума, решается с помощью дополнительного исследования.

Геометрическая интерпретация метода для $n = 2$ приведена на рис. 1.

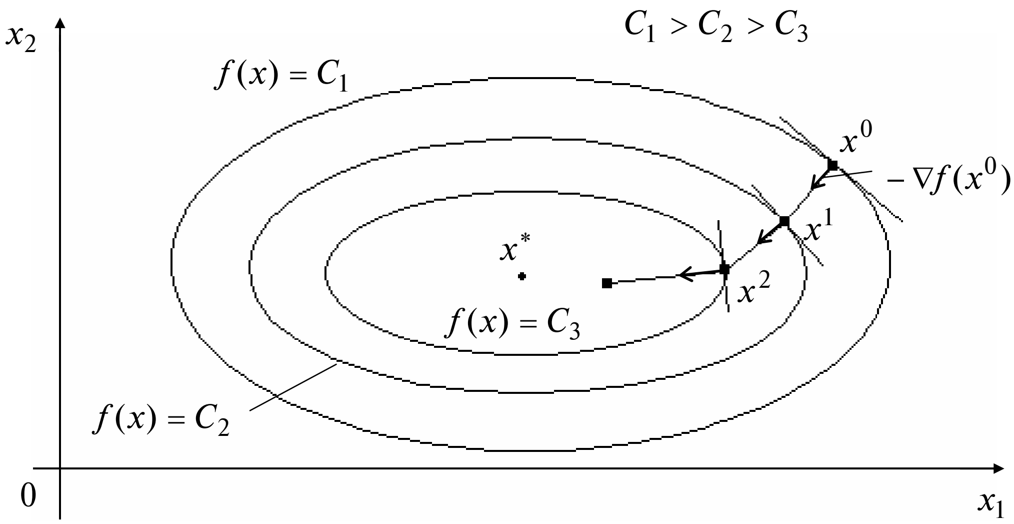

Рис. 1

З а м е ч а н и я.

1. Методы первого порядка при определенных условиях гарантируют сходимость последовательности $\left\{x^k\right\}$ к стационарной точке x^*, где $\nabla f\left(x^*\right) = 0$. Следовательно, найденная в результате применения метода точка x^* нуждается в дополнительном исследовании с целью ее классификации.

2. Условия окончания процесса поиска для большинства методов первого и второго порядков одни и те же (совпадают с применяемыми в методе А).

Б. МЕТОД НАИСКОРЕЙШЕГО ГРАДИЕНТНОГО СПУСКА

Стратегия поиска

Стратегия решения задачи состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу

$$x^{k+1} = x^k - t_k \nabla f\left(x^k\right),$$

где точка x^0 задается пользователем; величина шага t_k определяется для каждого значения *k* из условия

$$\varphi(t_k) = f\left(x^k - t_k \nabla f\left(x^k\right)\right) \to \min_{t_k}.$$

Эта задача может решаться с использованием необходимого условия минимума $\dfrac{d\varphi}{dt_k} = 0$ и последующей проверкой достаточного условия минимума $\dfrac{d^2\varphi}{dt_k{}^2} > 0$. Другой путь связан с использованием численных методов, когда ищется $\min\limits_{t_k \in [a,b]} \varphi(t_k) = \min\limits_{t_k \in [a,b]} f\left(x^k - t_k \nabla f\left(x^k\right)\right)$ с помощью методом одномерной минимизации.

Геометрическая интерпретация метода для $n = 2$ приведена на рис. 2.

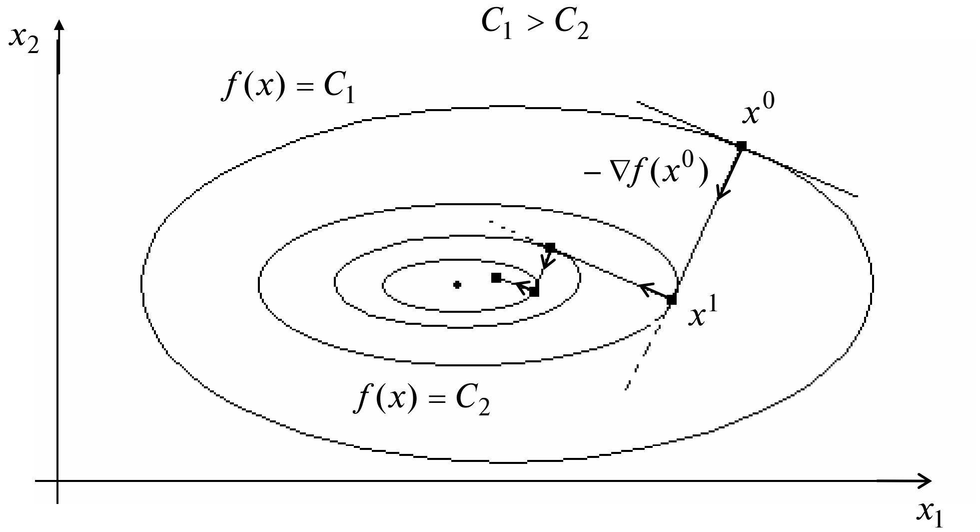

Рис. 2

В. МЕТОД ПОКООРДИНАТНОГО СПУСКА

Стратегия поиска

Стратегия решения задачи состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности $\left\{x^k\right\}$ вычисляются по циклам в соответствии с правилом

$$x^{jk+1} = x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}}\right)_{x = x^{jk}} \cdot e_{k+1},$$

где j – номер цикла вычислений; $j = 0,1,2,\ldots$; k – номер итерации внутри цикла, $k = 0,1,\ldots,n-1$; e_{k+1}, $k = 0,1,\ldots,n-1$ – единичный вектор, $(k+1)$-я проекция которого равна 1; точка x^{00} задается пользователем; величина шага t_k выбирается из условия

$$f\left(x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}}\right)_{x = x^{jk}} \cdot e_{k+1}\right) - f\left(x^{jk}\right) < 0 \text{ или } f\left(x^{jk+1}\right) - f\left(x^{jk}\right) < -\varepsilon \left\| \nabla f\left(x^{jk}\right) \right\|^2.$$

Если выбранное условие при текущем t_k не выполняется, шаг уменьшается вдвое и точка $x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}}\right)_{x = x^{jk}} \cdot e_{k+1}$ вычисляется заново. Легко видеть, что при фиксированном j за одну итерацию с номером k изменяется только одна проекция точки x^{jk}, имеющая номер $k+1$, а в течение всего цикла с номером j, т.е. начиная с $k=0$ и кончая $k=n-1$, изменяются все n проекций точки x^{j0}. После этого точке x^{jn} присваивается номер $x^{j+1,0}$ и она берется за начальную точку для вычислений в $(j+1)$-м цикле.

Полученные в результате вычислений точки могут быть записаны как элементы последовательности $\left\{x^l\right\}$, где $l=n\cdot j+k$ – порядковый номер точки, т.е. $\left\{x^l\right\}=\left\{x^0=x^{00}, x^1=x^{01},\ldots,x^n=x^{0n}=x^{10}, x^{n+1}=x^{11}, x^{n+2}=x^{12},\ldots\right\}$.

Геометрическая интерпретация метода для $n=2$ приведена на рис. 3.

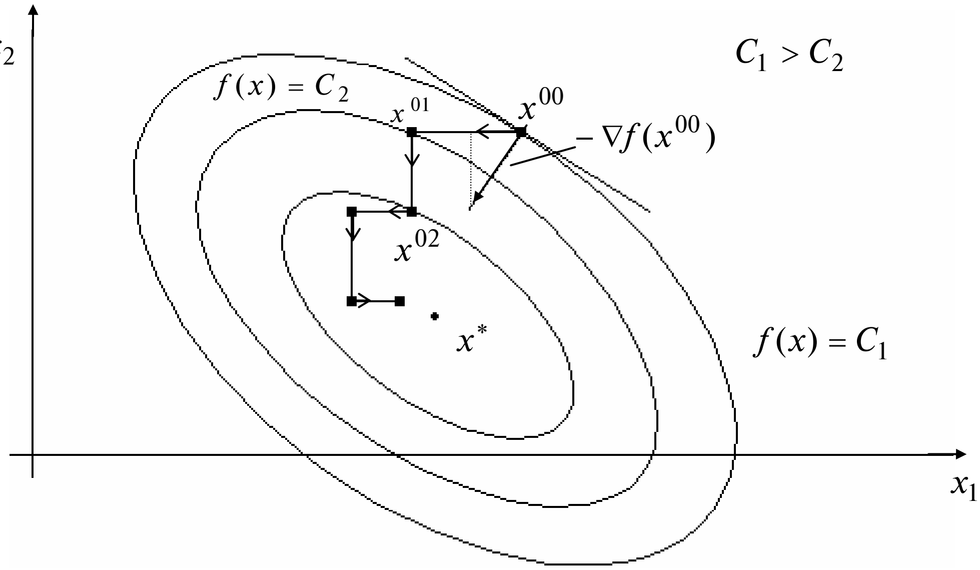

Рис. 3

Г. МЕТОД ГАУССА–ЗЕЙДЕЛЯ

Стратегия поиска

Стратегия метода Гаусса–Зейделя (Gauss–Seidel) состоит в построении последовательности точек $\left\{x^k\right\}$, $k=0,1,\ldots$, таких, что $f\left(x^{k+1}\right)<f\left(x^k\right)$, $k=0,1,\ldots$. Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу

$$x^{jk+1}=x^{jk}-t_k\left(\frac{\partial f(x)}{\partial x_{k+1}}\right)_{x=x^{jk}}\cdot e_{k+1},$$

где j – номер цикла вычислений, $j=0,1,2,\ldots$; k – номер итерации внутри цикла, $k=0,1,\ldots,n-1$; e_{k+1} – единичный вектор, $(k+1)$-я проекция которого равна 1; точка x^{00} задается пользователем, величина шага t_k выбирается из условия

$$\varphi(t_k)=f\left(x^{jk}-t_k\left(\frac{\partial f(x)}{\partial x_{k+1}}\right)_{x=x^{jk}}\cdot e_{k+1}\right)\to\min_{t_k}.$$

Данная задача является задачей одномерной минимизации функции $\varphi(t_k) = f\left(x^{jk} - t_k \left(\frac{\partial f(x)}{\partial x_{k+1}} \right)_{x = x^{jk}} \cdot e_{k+1} \right)$ и может быть решена либо с использованием условий $\frac{d\varphi}{dt_k} = 0$, $\frac{d^2\varphi}{dt_k^{\ 2}} > 0$, либо численно с использованием методов одномерной минимизации, как задача $\varphi(t_k) \to \min\limits_{t_k \in [a,b]}$.

При фиксированном j за одну итерацию с номером k изменяется только одна проекция точки x^{jk}, имеющая номер $k+1$, а в течение всего цикла с номером j, т.е. начиная с $k = 0$ и кончая $k = n-1$, изменяются все n проекций точки x^{j0}. После этого точке x^{jn} присваивается номер $x^{j+1,0}$ и она берется за начальную точку для вычислений в $(j+1)$-м цикле.

Полученные в результате вычислений точки могут быть записаны как элементы последовательности $\left\{x^l\right\}$, где $l = n \cdot j + k$ – порядковый номер точки, т.е. $\left\{x^l\right\} = \left\{x^0 = x^{00}, x^1 = x^{01}, \ldots, x^n = x^{0n} = x^{10}, x^{n+1} = x^{11}, x^{n+2} = x^{12}, \ldots\right\}$.

Геометрическая интерпретация метода для $n = 2$ приведена на рис. 4.

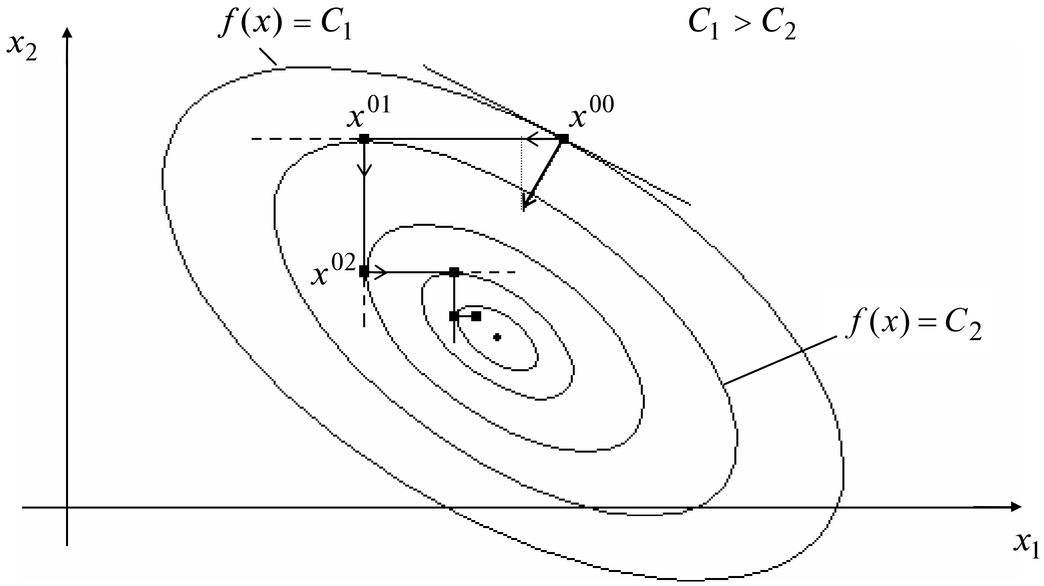

Рис. 4

Д. МЕТОД ФЛЕТЧЕРА–РИВСА

Стратегия поиска

Стратегия метода Флетчера–Ривса (Fletcher–Reeves) состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу:

$$x^{k+1} = x^k + t_k d^k, \ k = 0,1,\ldots;$$

$$d^0 = -\nabla f\left(x^0\right); \quad d^k = -\nabla f\left(x^k\right) + \beta_{k-1} d^{k-1}, \ k \geq 1;$$

$$\beta_{k-1} = \frac{\left\| \nabla f\left(x^k\right) \right\|^2}{\left\| \nabla f\left(x^{k-1}\right) \right\|^2}.$$

Точка x^0 задается пользователем, величина шага t_k определяется для каждого значения k из условия $\varphi(t_k) = f\left(x^k + t_k d^k\right) \to \min\limits_{t_k}$.

Решение задачи одномерной минимизации может осуществляться либо из условия $\dfrac{d\varphi}{dt_k} = 0$, $\dfrac{d^2\varphi}{dt_k{}^2} > 0$, либо численно с использованием методов одномерной минимизации, когда решается задача $\varphi(t_k) \to \min\limits_{t_k \in [a,b]}$.

Для квадратичных функций $f(x)$ с матрицей Гессе $H > 0$ метод Флетчера–Ривса сходится к точке минимума за число шагов, не превышающее n – размерность вектора x.

Для неквадратичных функций, как правило, используется *алгоритм Полака–Рибьера* (Polak–Ribiere), когда величина β_{k-1} вычисляется следующим образом:

$$\beta_{k-1} = \begin{cases} \dfrac{\left(\nabla f\left(x^k\right), \left[\nabla f\left(x^k\right) - \nabla f\left(x^{k-1}\right)\right]\right)}{\left\| \nabla f\left(x^{k-1}\right) \right\|^2}, & k \notin J, \\ 0, & k \in J, \end{cases}$$

где $J = \{0, n, 2n, \ldots\}$. В отличие от алгоритма Флетчера–Ривса алгоритм Полака–Рибьера предусматривает использование итерации наискорейшего градиентного спуска через каждые n шагов. Геометрическая интерпретация метода для $n = 2$ изображена на рис. 5.

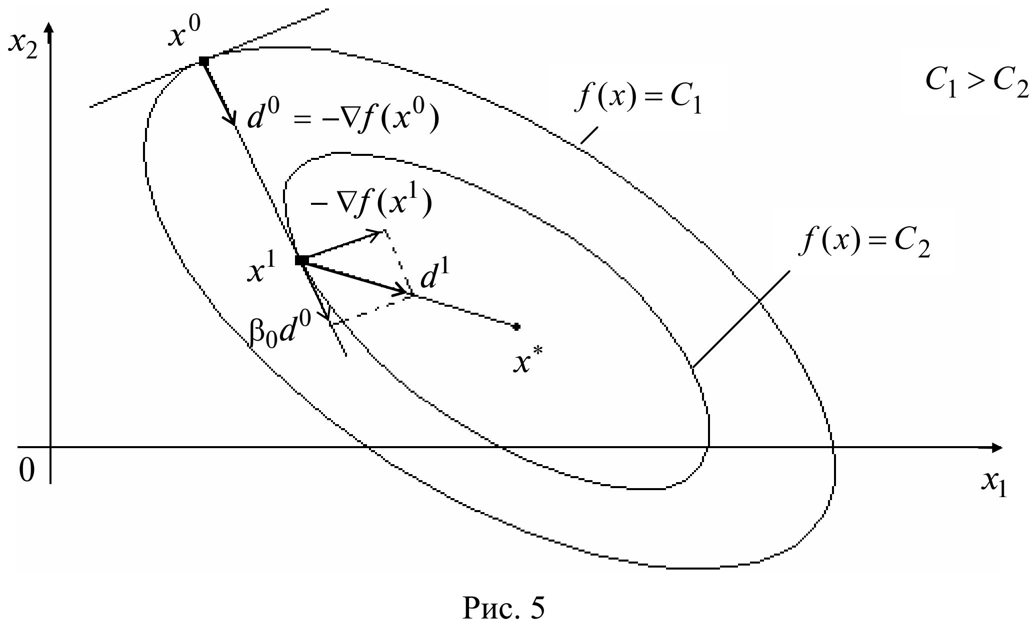

Рис. 5

МЕТОДЫ ВТОРОГО ПОРЯДКА

А. МЕТОД НЬЮТОНА

Стратегия поиска

Стратегия метода Ньютона (Newton) состоит в построении последовательности точек $\{x^k\}, k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности вычисляются по правилу

$$x^{k+1} = x^k - H^{-1}\left(x^k\right)\nabla f\left(x^k\right), \quad k = 0,1,\ldots ,$$

где x^0 – задается пользователем, а направление спуска d^k определяется для каждого значения k по формуле $d^k = -H^{-1}\left(x^k\right)\nabla f\left(x^k\right)$, величина шага $t_k = 1$.

З а м е ч а н и е. Для квадратичных функций при определенных условиях [1,2] метод сходится к стационарной точке за одну итерацию.

Б. МЕТОД НЬЮТОНА–РАФСОНА

Стратегия поиска

Стратегия метода Ньютона–Рафсона (Newton–Raphson) состоит в построении последовательности точек $\{x^k\}, k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Точки последовательности вычисляются по правилу

$$x^{k+1} = x^k - t_k\, H^{-1}\left(x^k\right)\nabla f\left(x^k\right), \quad k = 0,1,\ldots ,$$

где x^0 задается пользователем, а величина шага t_k определяется из условия

$$\varphi(t_k) = f\left(x^k - t_k\, H^{-1}\left(x^k\right)\nabla f\left(x^k\right)\right) \to \min_{t_k}.$$

Эта задача может решаться либо аналитически с использованием необходимого условия минимума $\dfrac{d\varphi}{dt_k} = 0$ и последующей проверкой достаточного условия $\dfrac{d^2\varphi}{d\,t_k^{\,2}} > 0$, либо численно как задача $\varphi(t_k) \to \min\limits_{t_k \in [a,b]}$, где интервал $[a,b]$ задается пользователем.

З а м е ч а н и е. Для уменьшения числа обращений матрицы Гессе применяется *упрощенный метод Ньютона*, в котором обращение осуществляется один раз – в начальной точке x^0:

$$x^{k+1} = x^k - t_k H^{-1}\left(x^0\right)\nabla f\left(x^k\right), k = 0, 1, \ldots \ .$$

В. МЕТОД МАРКВАРДТА

Стратегия поиска

Стратегия метода Марквардта (Marquardt) состоит в построении последовательности точек $\left\{x^k\right\}$, $k = 0,1,\ldots$, таких, что $f\left(x^{k+1}\right) < f\left(x^k\right)$, $k = 0,1,\ldots$. Применяется в случаях, когда на какой-либо итерации (итерациях) выполняется условие $\det H(x^k) \approx 0$, что приводит к значительным погрешностям вычисления обратной матрицы.

Точки последовательности $\left\{x^k\right\}$ вычисляются по правилу

$$x^{k+1} = x^k - \left[H\left(x^k\right) + \mu^k E\right]^{-1} \nabla f\left(x^k\right),\ k = 0,1,\ldots ,$$

где точка x^0 задается пользователем, E – единичная матрица, μ^k – последовательность положительных чисел, таких, что матрица $\left[H\left(x^k\right) + \mu^k E\right]^{-1}$ положительно определена.

Как правило, число μ^0 назначается как минимум на порядок больше, чем самый большой элемент матрицы $H\left(x^0\right)$, а в ряде стандартных программ полагается $\mu^0 = 10^4$. Если $f\left(x^k - \left(H\left(x^k\right) + \mu^k E\right)^{-1} \nabla f\left(x^k\right)\right) < f\left(x^k\right)$, то $\mu^{k+1} = \dfrac{\mu^k}{2}$. В противном случае $\mu^{k+1} = 2\mu^k$. Легко видеть, что алгоритм Марквардта в зависимости от величины μ^k на каждом шаге по своим свойствам приближается либо к алгоритму Ньютона, либо к алгоритму градиентного спуска.