Раздел 7. Метод Ньютона решения нелинейных уравнений в ЛНП

Лекция 15 Метод Ньютона решения нелинейных уравнений в ЛНП.

Мы переходим к рассмотрению нелинейный уравнений в ЛНП. Без ограничения общности уравнение можем записать в виде

$$F(x) = O$$

где $F : X \to Y$ – нелинейное отображение из ЛНП X в ЛНП Y. (В ЛНП можем всё перенести в одну сторону.)

Общего метода нет, каждую задачу надо рассматривать индивидуально но есть классы задач, для которых можно выработать единые подходы.

Первоначальное исследование задачи: аналитическое, качественное. Вопросы существования, единственности решения, если единственности нет – структура множества решений. Может быть и бесконечное множество решений (например, если отображение линейно, у него может быть нетривиальное ядро – линеал, для нелинейных отображений – гиперповерхность или, как говорят, многообразие). Аналитические методы, которые иногда позволяют найти решение явно, и численные, когда решение (решения) ищется приближённо.

Мы ограничимся случаем, когда множество решений конечно, расположение одного из них приблизительно известно. Если решение единственно, то иногда (не всегда) можно в качестве начального приближения взять любую точку. Например, в методе простой итерации, когда пространство банахово, а уравнение приводится к виду $x = G(x)$ (буква F у нас занята), причём оператор G в правой части – сжатие на всём пространстве. Вообще же выбор начального приближения важен: итерационная последовательность при одном выборе начальной точки может сходиться, при другом расходиться, причём если решений несколько, то и сходиться последовательность может к разным решениям в зависимости от начальной точки.

Итак, пусть решение отделено, то есть известна область пространства X, где решение существует и единственно, и выбрано начальное приближение. Метод Ньютона – итерационная процедура уточнения этого решения, построения последовательности, к нему сходящейся, а вопросы о единственности решения и поиске других решений в случае неединственности выносим за скобки. На самом деле метод Ньютона – это одна из модификаций метода простой итерации, но довольно специфическая модификация.

Будем предполагать, что отображение F достаточно гладкое (дифференцируемое по Фреше, а прочие требования уточним далее). Обозначим начальную точку x_0, причём $F(x_0) \neq O$, в противном случае задача решена. Нам же нужно найти точку $x^* \in X$, для которой $F(x^*) = O$.

Обозначим $h = x^* - x_0$, тогда $x^* = x_0 + h$, и тогда наше уравнение примет вид

$$F(x_0 + h) = O\,.$$

Воспользуемся известным нам выражением, связывающим приращение функции с приращением аргумента:

$$F(x_0 + h) - F(x_0) = F'(x_0)h + \omega(x_0, h)\,,$$

где $F'(x_0)$ – производная Фреше отображения F в точке x_0, ограниченный линейный оператор из X в Y, а $\|\omega(x_0, h)\| = o(\|h\|)$. Последняя оценка, строго говоря, ничего нам не даёт, поскольку описывает поведение поправочного слагаемого при $\|h\| \to 0$, в то время как у нас эта величина конечна. Тем не менее, предположим, что мы выбрали начальное приближение достаточно хорошо, так что величина $\|h\|$ мала настолько, чтоб последним слагаемым можно в главном приближении пренебречь. В результате получаем, с учётом $F(x_0 + h) = F(x^*) = O$, приближённое равенство

$$-F(x_0) \approx F'(x_0)h\,.$$

Фактически мы устроили линеаризацию зависимости $F(x)$ в окрестности точки x_0 и заменили приращение функции её дифференциалом Фреше. В результате вместо точного, но нелинейного уравнения получили приближённое, но линейное, и надеемся, что его решение не очень сильно отличается от точного: $h_0 \approx h$, где h_0 – решение линейного уравнения

$$F'(x_0)h_0 = -F(x_0)\,.$$

Для того, чтобы полученное уравнение было однозначно разрешимо, потребуем, чтобы у оператора $F'(x_0)$ существовал ограниченный обратный. Это, между прочим, накладывает некоторые ограничения на класс рассматриваемых задач. В частности, если пространства X и Y конечномерные, то оператор F' задаётся матрицей, так что необходимым условием существования обратной матрицы является совпадение размерностей пространств X и Y, в противном случае матрица не будет квадратной. А в случае, если размерности совпадают, мы приходим к требованию невырожденности матрицы, задающей оператор $F'(x_0)$. Если же пространства бесконечномерные, то требование существования ограниченного обратного у $F'(x_0)$ означает, помимо прочего, что отображение F является локально биективным, то есть взаимно-однозначно отображает некоторую окрестность точки $x_0 \in X$ на некоторую окрестность точки $F(x_0) \in Y$. Отметим, что мы требуем не просто обратимости, а непрерывной обратимости производной, поскольку если обратный оператор существует, но не является ограниченным, то мы сталкиваемся с некорректной задачей, в которой малому значению правой части может соответствовать сколь угодно большое по норме значение h_0 - и, более того, область определения такого оператора (т.е. область значений $F'(x_0)$) может и не содержать вектора $-F(x_0)$.

Итак, пусть ограниченный обратный существует, тогда мы находим вектор h_0:

$$h_0 = -(F'(x_0))^{-1}F(x_0)\,.$$

Вычислим сумму: $x_1 = x_0 + h_0 = x_0 - (F'(x_0))^{-1}F(x_0)$. Поскольку мы пренебрегли поправочными членами, $h_0 \neq h$, и поэтому $x_1 \neq x^*$. Тем не менее, есть основания надеяться, что x_1 будет ближе к искомому решению x^*, чем начальное приближение x_0. Далее мы можем вычислить значение $F(x_1)$ и $F'(x_1)$ и при условии непрерывной обратимости последнего оператора найти h_1 и следующее приближение x_2 и т.д. В результате мы получаем рекуррентную формулу

$$x_{n+1} = x_n - (F'(x_n))^{-1}F(x_n)\,,$$

связывающую два последовательных приближения. Эта формула и определяет итерационный процесс метода Ньютона.

Поскольку в методе Ньютона требуется последовательно обращать операторы $F'(x_n)$, разумно потребовать равномерной ограниченности обратных операторов либо во всём пространстве, либо хотя бы в окрестности искомой точки:

$$\left\|(F'(x_n))^{-1}\right\| \le m\,,$$

где m – некоторая константа. Это условие носит название условия регулярности.

Итерационную формулу мето Ньютона мы получили на основании "пальцевых"нестрогих, хотя и правдоподобных эвристических соображений. Теперь проанализируем, при каких же условиях на начальное приближение и на отображение F мы можем гарантировать сходимость итерационной последовательности. Отметим сразу, что если последовательность сходится, то её пределом будет решение уравнения: переходя к пределу и воспользовавшись непрерывностью операторов F и $(F')^{-1}$, мы получим:

$$x^* = x^* - (F'(x^*))^{-1}F(x^*)\,,$$

что эквивалентно $F(x^*) = O$.

Для анализа сходимости последовательности метода Ньютона предположим дополнительно, что производная обладает свойством липшецевости, т.е. что

$$\|F'(\hat{x}) - F'(\tilde{x})\| \le N_2\|\hat{x} - \tilde{x})\|\,.$$

В этом случае, как мы знаем, поправочный член, описывающий разность между приращением и дифференциалом функции, можно оценить:

$$\|F(x+h) - F(x) - F'(x)h\| = \|\omega(x,h)\| \le \frac{N_2}{2}\|h\|^2\,.$$

Преобразуем теперь формулу метода Ньютона, вычтя из левой и правой части x_*, а также вычтя из $F(x_n)$ равное нулю значение $F(x^*)$ и внеся минус в скобки:

$$x_{n+1} - x^* = (x_n - x^*) + (F'(x_n))^{-1}(F(x^*) - F(x_n))\,.$$

Теперь заметим, что

$$F(x^*) - F(x_n) = F'(x_n)(x^* - x_n) + \omega(x_n, x^* - x_n)\,,$$

и поэтому

$$(F'(x_n))^{-1}(F(x^*) - F(x_n)) = (x^* - x_n) + (F'(x_n))^{-1}\omega(x_n, x^* - x_n)\,,$$

а потому

$$x_{n+1} - x^* = (F'(x_n))^{-1}\omega(x_n, x^* - x_n)\,,$$

откуда

$$\|x_{n+1} - x^*\| \le m \cdot \frac{N_2}{2}\|x_n - x^*\|^2\,.$$

Умножив левую и правую часть полученного неравенства на $mN_2/2$, получим:

$$\frac{mN_2}{2}\|x_{n+1} - x^*\| \le \left(\frac{mN_2}{2}\|x_n - x^*\|\right)^2\,.$$

По индукции доказывается, что

$$\frac{mN_2}{2}\|x_n - x^*\| \le \left(\frac{mN_2}{2}\|x_0 - x^*\|\right)^{2^n}\,,$$

откуда

$$\|x_n - x^*\| \le \frac{2}{mN_2}\left(\frac{mN_2}{2}\|x_0 - x^*\|\right)^{2^n} = \left(\frac{mN_2}{2}\right)^{2^n-1}\|x_0 - x^*\|^{2^n}\,.$$

Если $(mN_2/2)\|x_0 - x^*\| < 1$ или, что то же самое,

$$\|x_0 - x^*\| < \frac{2}{mN_2}\,,$$

то полученная оценка гарантирует стремление $\|x_n - x^*\|$ к нулю и, тем самым, сходимость последовательность x_n к x^*. Если последнее неравенство не выполняется, то последовательность метода Ньютона может и не сходиться. Требования к выбору начального приближения тем жёстче, чем больше норма оператора $(F')^{-1}$ и чем больше константа N_2, описывающая погрешность линейной аппроксимации.

Метод Ньютона, как уже было сказано, является вариантом метода простой итерации для уравнения

$$x = G(x)\,,$$

где

$$G(x) = x - (F'(x))^{-1}F(x)\,,$$

очевидно, эквивалентного исходному. В то же время итерационная последовательность метода Ньютона сходится куда быстрее, чем типичная последовательность метода простой итерации. Чтобы понять, в чём тут причина, рассмотрим одномерный вариант метода Ньютона – метод касательных.

Метод касательных используется при решении нелинейных уравнений вида

$$f(x) = 0\,.$$

Итерационная формула в этом случае имеет вид

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}\,.$$

Производная Фреше в этом случае – обычная производная $f'(x)$, а обратный оператор $(F'(x_n))^{-1}$ – оператор умножения на $(f'(x_n))^{-1}$. В роли константы m выступает величина $\max|f'(x)|^{-1} = (\min|f'(x)|)^{-1}$, а $N_2 = \max|f''(x)|$ (если, конечно, функция f дважды дифференцируема).

Формула метода касательных – это формула метода простых итераций для уравнения

$$x = g(x)\,,$$

где

$$g(x) = x - \frac{f(x)}{f'(x)}\,.$$

Посмотрим, когда g является сжатием. Для этого вычислим производную:

$$g'(x) = 1 - \frac{f'^2(x) - f(x)f''(x)}{f'^2(x)} = \frac{f(x)f''(x)}{f'^2(x)}\,.$$

Как мы помним, достаточным условием того, что функция осуществляет сжатие, является неравенство

$$|g'(x)| \leq q < 1\,.$$

Если это неравенство выполняется всюду, то при любом выборе начального приближения итерационная последовательность сходится. В противном случае следует выбрать отрезок в окрестности корня, который переводится функцией g в себя, и на котором значение функции достаточно близко к нулю. Достаточным условием сжатия является выполнение неравенства

$$\max|f(x)| \cdot m^2 N_2 \leq q < 1\,,$$

при этом

$$|x - x^*| \leq m|f(x)| \leq \frac{q}{mN_2}\,.$$

Чем меньший отрезок, окружающий корень, мы возьмём, тем меньше на нём значения $|f(x)|$ и тем меньше, соответственно, можно выбрать значение q – то есть тем более быструю сходимость нам гарантирует теорема о неподвижной точке. Поэтому по мере того, как последовательность приближается к корню, уменьшается значение $|f(x)|$, и мы оказываемся в области со всё меньшими эффективными значениями q, что и обеспечивает увеличение скорости сходимости.

Откуда коэффициент 2? Из-за того, что одна из точек x^*, где $g'(x^*) = 0$. Действительно,

$$
\begin{aligned}
|x_{n+1} - x^*| &= |g(x_n) - g(x^*)| = |g'(\tilde{x})| \cdot |x_n - x^*| \approx \\
\approx \frac{|g'(x_n) + g'(x^*)|}{2}|x_n - x^*| &= \frac{|g'(x_n)|}{2}|x_n - x^*| = \\
= \frac{|f(x_n) f''(x_n)|}{2 f'^2(x_n)}|x_n - x^*| &\approx \left|\frac{f''(x_n)}{2f'(x_n)}\right| \cdot |x_n - x^*|^2 \,,
\end{aligned}
$$

поскольку

$$
f(x_n) = f(x_n) - f(x^*) \approx f'(x_n)(x_n - x^*)\,.
$$

Замечание. Область, в которой можно выбирать начальную точку метода Ньютона, шире, чем область, где отображение g является сжимающим. То есть расстояние меду образами произвольных двух точек этой области может оказаться больше, чем расстояние между прообразами, однако расстояние между образом любой из этих точек до точки x^* всё равно будет меньше, чем расстояние от прообраза до этой точки.