Lecture notes

Thursday Feb 6, 2020

1 Order of convergence

Definition 1.1 (From Süli & Mayers)**.** Suppose that

$$\lim_{k\to\infty} x_k = \xi.$$

If there exists a sequence $\{\epsilon_k\}$ of positive numbers converging to 0 and $\mu > 0$ such that

$$|x_k - \xi| \leq \epsilon_k,\ k = 0, 1, 2, \ldots, \qquad \text{and} \qquad \lim_{k\to\infty} \frac{\epsilon_{k+1}}{\epsilon_k^q} = \mu,$$

then we say that the sequence $\{x_j\}$ converges to ξ **with at least order** $q > 1$.

Furthermore, if the above assumptions hold with $|x_k - \xi| \leq \epsilon_k$ (instead of $\leq$), then the sequence is said to converge to ξ **with order** q. If $q = 2$ then the sequence is said to converge quadratically.

Example 1.1. We proved in class that Newton's method converges quadratically, and that in fact,

$$\lim_{k\to\infty} \frac{|x_{k+1} - \xi|}{|x_k - \xi|^2} = \frac{1}{2}\left|\frac{f''(\xi)}{f'(\xi)}\right|.$$

Example 1.2. Recall that the secant method for solving $f(x) = 0$ is given by:

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} f(x_k).$$

The fraction above is a *finite difference* approximation to the value $1/f'(x_)$, which is used in Newton's method. One can show that the secant method converges with order in between 1 and 2, in fact, $q = (1 + \sqrt{5})/2$ for the secant method. (This is a homework exercise.)

2 Multi-dimensional Newton's method

Newton's method can also be used to solve *systems* of nonlinear equations. The general two-variable case is given as:

$$\begin{aligned} f_1(x_1, x_2) &= 0 \\ f_2(x_1, x_2) &= 0 \end{aligned}$$

This system can be written concisely in vector notation simply as $\mathbf{f}(\mathbf{x}) = \mathbf{0}$.

To begin, let's first recall the form of multivariate Taylor series expansions. In the case of a *scalar* function f of two variables expanded about the point $\mathbf{y} = (y_1, y_2)$:

$$\begin{aligned} f(\mathbf{x}) = f(\mathbf{y}) &+ \frac{\partial f}{\partial x_1}(\mathbf{y})(x_1 - y_1) + \frac{\partial f}{\partial x_2}(\mathbf{y})(x_2 - y_2) \\ &+ \frac{1}{2!}\left(\frac{\partial^2 f}{\partial x_1^2}(\mathbf{y})(x_1 - y_1)^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{y})(x_1 - y_1)(x_2 - y_2)\right. \\ &\left. + \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{y})(x_1 - y_1)(x_2 - y_2) + \frac{\partial^2 f}{\partial x_2^2}(\mathbf{y})(x_2 - y_2)^2 + \right) + \dots \end{aligned} \tag{1}$$

For convenience, we will suppress the evaluation point $\mathbf{y}$, it will be implied. The above expression can be written more succinctly as

$$\begin{aligned} f(\mathbf{x}) &= f + \begin{pmatrix} f_{x_1} & f_{x_2} \end{pmatrix} (\mathbf{x} - \mathbf{y}) + (\mathbf{x} - \mathbf{y})^T \begin{pmatrix} f_{x_1 x_1} & f_{x_1 x_2} \\ f_{x_2 x_1} & f_{x_2 x_2} \end{pmatrix} (\mathbf{x} - \mathbf{y}) + \dots \\ &= f + Df(\mathbf{x} - \mathbf{y}) + (\mathbf{x} - \mathbf{y})^T D^T Df(\mathbf{x} - \mathbf{y}) + \dots \end{aligned}$$

The 2×2 matrix $D^T Df$ is usually called the Hessian of f.

Now, unfortunately, we need to do a related calculation for a *vector-valued* function, $\mathbf{f} = \mathbf{f}(\mathbf{x})$ instead of a scalar-valued one. In this case, the multivariable Taylor series is just computed for each component of the vector f, which is just a vector of functions:

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \end{pmatrix}$$

In this case, the first two terms in the vector-valued multivariate Taylor series are:

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{y}) + \mathbf{J}(\mathbf{x} - \mathbf{y}) + \dots$$

where the matrix J is known as the Jacobian matrix, and is given by:

$$\mathbf{J} = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{pmatrix}$$

The higher order terms in the above Taylor expansion formally involve tensors, which will not discuss. With the above expansion, we can derive Newton's method in the vector-valued case. Recall what Newton's method in one dimension does: it approximates the nonlinear function f by a linear one using the tangent line, and then finds the root of the tangent line. This root is used as an approximation to the root, and a new tangent line is formed, and the process repeats.

If we were to truncate the above infinite expansion, we would have:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{y}) + \mathbf{J}(\mathbf{x} - \mathbf{y}).$$

Furthermore, if we denote the root of $\mathbf{f}$ by $\mathbf{z}$, then we have that

$$\mathbf{f}(\mathbf{z}) = \mathbf{0} \approx \mathbf{f}(\mathbf{y}) + \mathbf{J}(\mathbf{y})(\mathbf{z} - \mathbf{y}),$$

where the evaluation point of J is made explicit. Solving the above approximate equation for $\mathbf{z}$ we have that:

$$\mathbf{z} \approx \mathbf{y} - \mathbf{J}(\mathbf{y})^{-1}\mathbf{f}(\mathbf{y}).$$

Therefore, starting with an initial guess for the root $\mathbf{x}_0$, we have that the multidimensional Newton's method is:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{J}(\mathbf{x}_k)^{-1}\mathbf{f}(\mathbf{x}_k).$$

Convergence for the multidimensional Newton's method is also quadratic (under some assumptions similar to those in the one-dimensional case, which we will not detail). In this case, we must replace the absolute errors with norms:

$$\|\mathbf{x}_{k+1} - \mathbf{z}\| \approx A\|\mathbf{x}_k - \mathbf{z}\|$$

where A is some constant, and $\|\cdot\|$ denotes the usual ℓ_2 Euclidean norm:

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

The above expressions straightforwardly generalize to the n-dimensional case (as opposed to the 2-dimensional case that we worked out explicitly).