

LECTURE 23: NONLINEAR SCALAR OPERATOR EQUATIONS

1. Root-finding algorithms

The null space of a linear mapping represented through matrix $A \in \mathbb{C}^{m \times n}$ is defined as $N(A) = \{x | Ax = 0\}$, the set of all points that have a null image through the mapping. The null space is a vector subspace of the domain of the linear mapping. A first step in the study of nonlinear mappings is to consider the generalization of the concept of a null set, starting with the simplest case,

$$f(x) = 0 \quad (1)$$

where $f: \mathbb{R} \rightarrow \mathbb{R}$, $f \in C^p(\mathbb{R})$, $p \geq 0$, i.e., f has p continuous derivatives. It is assumed that a closed form analytical solution is not available, and algorithms are sought to construct an approximating sequence $\{x_n\}_{n \in \mathbb{N}}$ whose limit is a root of (1). The general approach is to replace (1) with

$$g_n(x) = 0, \quad (2)$$

where g_n is some approximation of f , and x_n the root of (2) can be easily determined.

1.1. First-degree polynomial approximants

Secant method. Consider $g(x) = ax + b$ (a linear function, but not a linear mapping for $b \neq 0$), an approximant of f based upon data $\{(x_{n-2}, f_{n-2} = f(x_{n-2})), (x_{n-1}, f_{n-1} = f(x_{n-1}))\}$, given in Newton interpolant form by

$$g_n(x) = f_{n-2} + \frac{f_{n-1} - f_{n-2}}{x_{n-1} - x_{n-2}}(x - x_{n-2}). \quad (3)$$

The solution of (3) is

$$x_n = x_{n-2} - \frac{f_{n-2}}{f_{n-1} - f_{n-2}}(x_{n-1} - x_{n-2}) = \frac{x_{n-2}f_{n-1} - x_{n-1}f_{n-2}}{f_{n-1} - f_{n-2}},$$

an iteration known as the secant method. The error in root approximation is

$$e_n = x_n - x = e_{n-2} - \frac{f_{n-2}}{f_{n-1} - f_{n-2}}(e_{n-1} - e_{n-2}),$$

and can be estimated by Taylor series expansions around the root x for which $f(x) = 0$,

$$f_{n-k} = f(x_{n-k}) = f'(x)(x_{n-k} - x) + \frac{1}{2}f''(x)(x_{n-k} - x)^2 + \dots = f' e_{n-k} + \frac{1}{2}f'' e_{n-k}^2 + \dots,$$

where derivatives f', f'' are assumed to be evaluated at x . In the result

$$e_n = e_{n-2} - \frac{f' e_{n-2} + \frac{1}{2}f'' e_{n-2}^2 + \dots}{f' \cdot (e_{n-1} - e_{n-2}) + \frac{1}{2}f'' \cdot (e_{n-1}^2 - e_{n-2}^2) + \dots} (e_{n-1} - e_{n-2}) = e_{n-2} \left[1 - \frac{f' + \frac{1}{2}f'' \cdot e_{n-2} + \dots}{f' + \frac{1}{2}f'' \cdot (e_{n-1} + e_{n-2}) + \dots} \right],$$

assuming $f'(x) \neq 0$, (i.e., x is a simple root) gives

$$e_n = e_{n-2} \left[1 - \frac{1 + c \cdot e_{n-2} + \dots}{1 + c \cdot (e_{n-1} + e_{n-2}) + \dots} \right], c = \frac{1}{2}(f''/f').$$

For small errors, to first order the above can be written as

$$e_n = e_{n-2} [1 - (1 + c \cdot e_{n-2})(1 - c \cdot (e_{n-1} + e_{n-2}))] = c e_{n-1} e_{n-2}.$$

Assuming p -order convergence of e_n ,

$$|e_n| \sim A |e_{n-1}|^p,$$

leads to

$$A^{p+1} |e_{n-2}|^{p^2} \sim c A |e_{n-2}|^{p+1} \Rightarrow |e_{n-2}|^{p^2-p-1} \sim c A^{-p}.$$

Since c, A are finite while $e_n \rightarrow 0$, the above asymptotic relation can only be satisfied if

$$p^2 - p - 1 = 0 \Rightarrow p = \frac{1 + \sqrt{5}}{2} \cong 1.62,$$

hence the secant method exhibits superlinear, but subquadratic convergence.

Newton-Raphson method. A different linear approximant arises from the Hermite interpolant based on data

$$\{(x_{n-1}, f_{n-1} = f(x_{n-1}), f'_{n-1} = f'(x_{n-1}))\},$$

which is given in Newton form as

$$g_n(x) = f_{n-1} + f'_{n-1} \cdot (x - x_{n-1}),$$

with root

$$x_n = x_{n-1} - \frac{f_{n-1}}{f'_{n-1}}, \quad (4)$$

an iteration known as the Newton-Raphson method. The error is given by

$$e_n = x_n - x = e_{n-1} - \frac{f_{n-1}}{f'_{n-1}}. \quad (5)$$

Taylor series expansion around the root gives for small e_{n-1} ,

$$e_n = e_{n-1} - \frac{f' \cdot e_{n-1} + \frac{1}{2} f'' \cdot e_{n-1}^2 + \dots}{f' + f'' e_{n-1} + \dots} = e_{n-1} \left[1 - \frac{1 + c e_{n-1} + \dots}{1 + 2c e_{n-1} + \dots} \right] \approx e_{n-1} [1 - (1 + c e_{n-1})(1 - 2c e_{n-1})].$$

The resulting expression

$$e_n \approx c e_{n-1}^2 = \frac{1}{2} \frac{f''}{f'} e_{n-1}^2, \quad (6)$$

states quadratic convergence for Newton's method. This faster convergence than the secant method requires however knowledge of the derivative, and the computational expense of evaluating it.

The above estimate assumes convergence of $\{x_n\}_{n \in \mathbb{N}}$, but this is not guaranteed in general. Newton's method requires an accurate initial approximation x_0 , within a neighborhood of the root in which f is increasing, $f' > 0$, and convex, $f'' > 0$. Equivalently, since roots of f are also roots of $-f$, Newton's method converges when $f', f'' < 0$. In both cases (6) in the prior iteration states that $e_{n-1} = x_{n-1} - r > 0$, hence $x_{n-1} > r$. Since f is increasing $f(x_{n-1}) > f(r) = 0$, hence (5) implies $e_n < e_{n-1}$. Thus the sequence $\{e_n\}_{n \in \mathbb{N}}$ is decreasing and bounded below by zero, hence $\lim_{n \rightarrow \infty} e_n = 0$, and Newton's method converges.

1.2. Second-degree polynomial approximants

An immediate extension of the above approach is to increase the accuracy of the approximant by seeking a higher-degree polynomial interpolant. The expense of the resulting algorithm increases rapidly though, and in practice linear and quadratic approximants are the most widely used. Consider the Hermite interpolant based on data

$$\{(x_{n-1}, f_{n-1} = f(x_{n-1}), f'_{n-1} = f'(x_{n-1}), f''_{n-1} = f''(x_{n-1}))\},$$

given in Newton form as

$$g_n(x) = f_{n-1} + f'_{n-1} \cdot (x - x_{n-1}) + \frac{1}{2} f''_{n-1} \cdot (x - x_{n-1})^2 = C + B s + A s^2,$$

with roots

$$x_n = x_{n-1} + \frac{-f'_{n-1} \pm \sqrt{(f'_{n-1})^2 - 2 f_{n-1} f''_{n-1}}}{f''_{n-1}}.$$

The above exhibits the difficulties arising in higher-order interpolants. The iteration requires evaluation of a square root, and checking for a positive discriminant.

Halley's method. Algebraic manipulations can avoid the appearance of radicals in a root-finding iteration. As an example, Halley's method

$$x_n = x_{n-1} - \frac{2 f_{n-1} f'_{n-1}}{2 (f'_{n-1})^2 - f_{n-1} f''_{n-1}},$$

exhibits cubic convergence.

2. Composite approximations

The secant iteration

$$x_n = x_{n-2} - \frac{f_{n-2}}{f_{n-1} - f_{n-2}}(x_{n-1} - x_{n-2}) = x_{n-2} - \frac{f_{n-2}}{\frac{f_{n-1} - f_{n-2}}{x_{n-1} - x_{n-2}}},$$

in the limit of $x_{n-2} \rightarrow x_{n-1}$ recovers Newton's method

$$x_n = x_{n-1} - \frac{f_{n-1}}{f'_{n-1}}.$$

This suggests seeking advantageous approximations of the derivative

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

based upon some step-size sequence $\{h_n\}$. Since $f(x_n) \rightarrow 0$, the choice $h_{n-1} = f(x_{n-1})$ suggests itself, leading to Steffensen's method

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

Steffensen's method exhibits quadratic convergence, just like Newton's method, but does not require knowledge of the derivative. The higher order by comparison to the secant method is a direct result of the derivative approximation

$$f'(x_{n-1}) \cong \frac{f(x_{n-1} + f(x_{n-1})) - f(x_{n-1})}{f(x_{n-1})},$$

which, remarkably, utilizes a composite approximation

$$f(x_{n-1} + f(x_{n-1})) = (f \circ (1 + f))(x_{n-1}).$$

Such composite techniques are a prominent feature of various nonlinear approximations such as a k -layer deep neural network $f(\mathbf{x}) = (I_k \circ I_{k-1} \circ \dots \circ I_1)(\mathbf{x})$.

3. Fixed-point iteration

The above iterative sequences have the form

$$x_n = F(x_{n-1}),$$

and the root is a fixed point of the iteration

$$x = F(x).$$

For example, in Newton's method

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

and indeed at a root $x = F(x)$. Characterization of mappings F that lead to convergent approximation sequences is of interest and leads to the following definition and theorem.

DEFINITION. A function $F: [a, b] \rightarrow [a, b]$ is said to be a *contractive mapping* if $\forall x, y \in [a, b]$ there exists $c \in (0, 1)$ such that

$$|F(x) - F(y)| \leq c|x - y|.$$

THEOREM. (*Contractive Mapping theorem*). If $F: [a, b] \rightarrow [a, b]$ is a contractive mapping then F has a unique fixed point $x \in [a, b]$, $x = F(x)$.

The fixed point theorem is an entry point to the study of non-additive approximation sequences.

Example 1. The sequence

$$x_1 = \sqrt{p}, x_2 = \sqrt{p + \sqrt{p}}, \dots \quad (p > 0)$$

is expressed recursively as

$$x_{n+1} = \sqrt{p + x_n},$$

and has the limit

$$x = \sqrt{p + \sqrt{p + \sqrt{p + \dots}}},$$

that is the fixed point of F ,

$$x = F(x) = \sqrt{p + x} = \frac{1 + \sqrt{1 + p}}{2}.$$

Over the interval $[0, p + 1]$, F is a contraction since

$$F'(x) = \frac{1}{2\sqrt{p + x}} \leq \frac{1}{2\sqrt{p}} < 1.$$

Example 2. The sequence

$$x_1 = \frac{1}{p}, x_2 = \frac{1}{p + \frac{1}{p}}, \dots \quad (p > 0)$$

is expressed recursively as

$$x_{n+1} = \frac{1}{p + x_n},$$

and has the limit

$$x = \frac{1}{p + \frac{1}{p + \dots}},$$

that is the fixed point of F ,

$$x = F(x) = \frac{1}{p + x} = \frac{-p + \sqrt{p^2 + 1}}{2}.$$

Over the interval $[0, 1]$, F is a contraction since

$$|F'(x)| = \frac{1}{(p + x)^2} \leq \frac{1}{p^2} < 1.$$