0.1 Numerical methods

0.1.1 | Errors

Floating-point representation

Theorem 1.1. Let $b \in \mathbb{N}$, $b \geq 2$. Any real number $x \in \mathbb{R}$ can be represented of the form

$$x = s\left(\sum_{i=1}^{\infty} \alpha_i b^{-i}\right) b^q,$$

where $s \in \{-1, 1\}$, $q \in \mathbb{Z}$ and $\alpha_i \in \{0, 1, \ldots, b-1\}$. Moreover, this representation is unique if $\alpha_1 \neq 0$ and $\forall i_0 \in \mathbb{N}$, $\exists i \geq i_0 : \alpha_i \neq b-1$. We will write

$$x = s(0.\alpha_1\alpha_2\cdots)_b b^q,$$

where the subscript b in the parenthesis indicates that the number $0.\alpha_1\alpha_2\alpha_3\cdots$ is in base b.

Definition 1.2 (Floating-point representation). Let x be a real number. Then the floating-point representation of x is

$$x = s\left(\sum_{i=1}^{t} \alpha_i b^{-i}\right) b^q.$$

Here s is called the *sign*; $\sum_{i=1}^{t} \alpha_i b^{-i}$, the *significant* or *mantissa*, and q, the *exponent*, limited to a prefixed range $q_{\min} \leq q \leq q_{\max}$. So, the floating-point representation of x is

$$x = smb^q = s(0.\alpha_1\alpha_2\cdots\alpha_t)_b b^q.$$

Finally we say a floating-point number is *normalized* if $\alpha_1 \neq 0$.

Definition 1.3. Let $x \in \mathbb{R}$ be such that $x = s(0.\alpha_1\alpha_2\cdots)_b b^q$ with $q_{\min} \leq q \leq q_{\max}$. We say the *floating-point representation by truncation of* x is

$$fl_T(x) = s(0.\alpha_1\alpha_2\cdots\alpha_t)_b b^q.$$

We say the *floating-point representation by rounding of* x is

$$fl_R(x) = \\ = \begin{cases} s(0.\alpha_1\cdots\alpha_t)_b b^q & \text{if } 0 \leq \alpha_{t+1} < \dfrac{b}{2} \\ s(0.\alpha_1\cdots\alpha_{t-1}(\alpha_t+1))_b b^q & \text{if } \dfrac{b}{2} \leq \alpha_{t+1} \leq b-1. \end{cases}$$

Definition 1.4. Given a value $x \in \mathbb{R}$ and an approximation $\tilde{x}$ of x, the *absolute error* is

$$\Delta x := |x - \tilde{x}|.$$

If $x \neq 0$, the *relative error* is

$$\delta x := \frac{|x - \tilde{x}|}{x}.$$

If x is unknown, we take

$$\delta x \approx \frac{|x - \tilde{x}|}{\tilde{x}}.$$

Definition 1.5. Let $\tilde{x}$ be an approximation of x. If $\Delta x \leq \frac{1}{2}10^{-t}$, we say $\tilde{x}$ *has t correct decimal digits*. If $x = sm10^q$ with $0.1 \leq m < 1$, $\tilde{x} = s\tilde{m}10^q$ and

$$u := \max\{i \in \mathbb{Z} : |m - \tilde{m}| \leq \frac{1}{2}10^{-i}\},$$

then we say that $\tilde{x}$ *has u significant digits.*

Proposition 1.6. Let $x \in \mathbb{R}$ be such that $x = s(0.\alpha_1\alpha_2\cdots)_b b^q$ with $\alpha_1 \neq 0$ and $q_{\min} \leq q \leq q_{\max}$. Then, its floating-point representation in base b and with t digits satisfy:

$$|fl_T(x) - x| \leq b^{q-t}, \qquad |fl_R(x) - x| \leq \frac{1}{2}b^{q-t}.$$
$$\left|\frac{fl_T(x) - x}{x}\right| \leq b^{1-t}, \qquad \left|\frac{fl_R(x) - x}{x}\right| \leq \frac{1}{2}b^{1-t}.$$

Definition 1.7. The *machine epsilon* ϵ is defined as

$$\epsilon := \min\{\varepsilon > 0 : fl(1 + \varepsilon) \neq 1\}.$$

Proposition 1.8. For a machine working by truncation, $\epsilon = b^{1-t}$. For a machine working by rounding, $\epsilon = \frac{1}{2}b^{1-t}$.

Propagation of errors

Proposition 1.9 (Propagation of absolute errors). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class $\mathcal{C}^2$. If Δx_j is the absolute error of the variable x_j and $\Delta f(x)$ is the absolute error of the function f evaluated at the point $x = (x_1, \ldots, x_n)$, we have

$$|\Delta f(x)| \lesssim \sum_{j=1}^{n} \left|\frac{\partial f}{\partial x_j}(x)\right| |\Delta x_j|^{1}.$$

The coefficients $\left|\frac{\partial f}{\partial x_j}(x)\right|$ are called *absolute condition numbers of the problem.*

Proposition 1.10 (Propagation of relative errors). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class $\mathcal{C}^2$. If δx_j is the relative error of the variable x_j and $\delta f(x)$ is the relative error of the function f evaluated at the point $x = (x_1, \ldots, x_n)$, we have

$$|\delta f(x)| \lesssim \sum_{j=1}^{n} \frac{\left|\frac{\partial f}{\partial x_j}(x)\right| |x_j|}{|f(x)|} |\delta x_j|.$$

The coefficients $\frac{\left|\frac{\partial f}{\partial x_j}(x)\right| |x_j|}{|f(x)|}$ are called *relative condition numbers of the problem.*

[1] The symbol $\lesssim$ means that we are omitting terms of order $\Delta x_j \Delta x_k$ and higher.

Numerical stability of algorithms

Definition 1.11. An algorithm is said to be *numerically stable* if errors in the input lessen in significance as the algorithm executes, having little effect on the final output. On the other hand, an algorithm is said to be *numerically unstable* if errors in the input cause a considerably larger error in the final output.

Definition 1.12. A problem with a low condition number is said to be *well-conditioned*. Conversely, a problem with a high condition number is said to be *ill-conditioned*.

0.1.2 | Zeros of functions

Definition 1.13. Let $f:\mathbb{R}\to\mathbb{R}$ be a function. We say α is a *zero* or a *solution to the equation* $f(x)=0$ if $f(\alpha)=0$.

Definition 1.14. Let $f:\mathbb{R}\to\mathbb{R}$ be a sufficiently differentiable function. We say α is a *zero of multiplicity* $m\in\mathbb{N}$ if

$$f(\alpha)=f'(\alpha)=\cdots=f^{(m-1)}(\alpha)=0\quad\text{and}\quad f^{(m)}(\alpha)\neq 0.$$

If $m=1$, the zero is called *simple*; if $m=2$, *double*; if $m=3$, *triple*...

Root-finding methods

For the following methods consider a continuous function $f:I\subseteq\mathbb{R}\to\mathbb{R}$ with an unknown zero $\alpha\in I$. Given $\varepsilon>0$, we want to approximate α with $\tilde{\alpha}$ such that $|\alpha-\tilde{\alpha}|<\varepsilon$.

Theorem 1.15 (Bisection method). Suppose $I=[a_0,b_0]$. For each step $n\geq 0$ of the algorithm we will approximate α by

$$c_n=\frac{a_n+b_n}{2}.$$

If $f(c_n)=0$ we are done. If not, let

$$[a_{n+1},b_{n+1}]=\begin{cases}[a_n,c_n] & \text{if}\quad f(a_n)f(c_n)<0,\\ [c_n,b_n] & \text{if}\quad f(a_n)f(c_n)>0.\end{cases}$$

and iterate the process again[2]. Observe the length of the interval $[a_n,b_n]$ is $\frac{b_0-a_0}{2^n}$ and therefore:

$$|\alpha-c_n|<\frac{b_0-a_0}{2^{n+1}}<\varepsilon\iff n>\frac{\log\left(\frac{b_0-a_0}{\varepsilon}\right)}{\log 2}-1.$$

Theorem 1.16 (*Regula falsi* method). Suppose $I=[a_0,b_0]$. For each step $n\geq 0$ of the algorithm we will approximate α by

$$c_n=b_n-f(b_n)\frac{b_n-a_n}{f(b_n)-f(a_n)}=\frac{a_nf(b_n)-b_nf(a_n)}{f(b_n)-f(a_n)}.$$

If $f(c_n)=0$ we are done. If not, let

$$[a_{n+1},b_{n+1}]=\begin{cases}[a_n,c_n] & \text{if}\quad f(a_n)f(c_n)<0,\\ [c_n,b_n] & \text{if}\quad f(a_n)f(c_n)>0,\end{cases}$$

and iterate the process again.

Theorem 1.17 (Secant method). Suppose $I=\mathbb{R}$ and that we have two different initial approximations x_0, x_1. Then for each step $n\geq 0$ of the algorithm we obtain a new approximation x_{n+2}, given by:

$$x_{n+2}=x_{n+1}-f(x_{n+1})\frac{x_{n+1}-x_n}{f(x_{n+1})-f(x_n)}.$$

Theorem 1.18 (Newton-Raphson method). Suppose $I=\mathbb{R}$, $f\in\mathcal{C}^1$ and that we have an initial approximation x_0. Then for each step $n\geq 0$ we obtain a new approximation x_{n+1}, given by:

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

Theorem 1.19 (Newton-Raphson modified method). Suppose $I=\mathbb{R}$, $f\in\mathcal{C}^1$ and that we have an initial approximation x_0 of a zero α of multiplicity m. Then for each step $n\geq 0$ we obtain a new approximation x_{n+1}, given by:

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

Theorem 1.20 (Chebyshev method). Suppose $I=\mathbb{R}$, $f\in\mathcal{C}^2$ and that we have an initial approximation x_0. Then for each step $n\geq 0$ we obtain a new approximation x_{n+1}, given by:

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)}-\frac{1}{2}\frac{[f(x_n)]^2f''(x_n)}{[f'(x_n)]^3}.$$

Fixed-point iterations

Definition 1.21. Let $g:[a,b]\to[a,b]\subset\mathbb{R}$ be a function. A point $\alpha\in[a,b]$ is *n-periodic* if $g^n(\alpha)=\alpha$ and $g^j(\alpha)\neq\alpha$ for $j=1,\ldots,n-1$[3].

Theorem 1.22 (Fixed-point theorem). Let (M,d) be a complete metric space and $g:M\to M$ be a contraction[4]. Then g has a unique fixed point $\alpha\in M$ and for every $x_0\in M$,

$$\lim_{n\to\infty}x_n=\alpha,\quad\text{where } x_n=g(x_{n-1})\quad\forall n\in\mathbb{N}.$$

Proposition 1.23. Let (M,d) be a metric space and $g:M\to M$ be a contraction of constant k. Then if we want to approximate a fixed point α by the iteration $x_n=g(x_{n-1})$, we have:

$$d(x_n,\alpha)\leq\frac{k^n}{1-k}d(x_1,x_0)\qquad\text{(a priori estimation)}$$

$$d(x_n,\alpha)\leq\frac{k}{1-k}d(x_n,x_{n-1})\qquad\text{(a posteriori estimation)}$$

Corollary 1.24. Let $g:\mathbb{R}\to\mathbb{R}$ be a function of class $\mathcal{C}^1$. Suppose α is a fixed point of g and $|g'(\alpha)|<1$. Then, there exists $\varepsilon>0$ and $I_\varepsilon:=[\alpha-\varepsilon,\alpha+\varepsilon]$ such that $g(I_\varepsilon)\subseteq I_\varepsilon$ and g is a contraction on I_ε. In particular, if $x_0\in I_\varepsilon$, the iteration $x_{n+1}=g(x_n)$ converges to α.

[2] Note that bisection method only works for zeros of odd multiplicity.
[3] Note that 1-periodic points are the fixed points of f.
[4] Remember definitions **??**, **??** and **??**.

Definition 1.25. Let $g : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^1$ and α be a fixed point of g. We say α is an *attractor fixed point* if $|g'(\alpha)| < 1$. In this case, any iteration $x_{n+1} = g(x_n)$ in I_ε converges to α. If $|g'(\alpha)| > 1$, we say α is a *repulsor fixed point*. In this case, $\forall x_0 \in I_\varepsilon$ the iteration $x_{n+1} = g(x_n)$ doesn't converge to α.

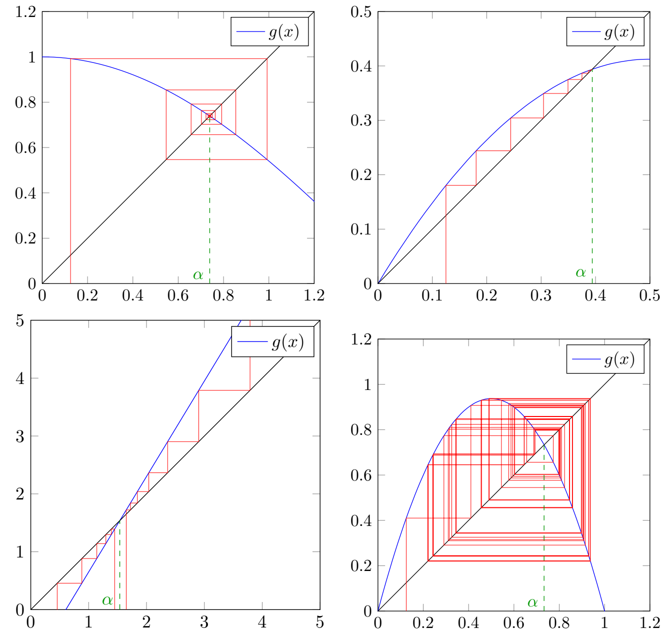

Figure 1: Cobweb diagrams. In the figures at the top, α is a attractor point, that is, $|g'(\alpha)| < 1$. More precisely, the figure at the top left occurs when $-1 < g'(\alpha) \le 0$ and the figure at the top right when $0 \le g'(\alpha) < 1$. In the figure at bottom left, α is a repulsor point. Finally, in the figure at bottom right the iteration $x_{n+1} = g(x_n)$ has no limit. It is said that to have a *chaotic behavior*.

Order of convergence

Definition 1.26 (Order of convergence). Let (x_n) be a sequence of real numbers that converges to $\alpha \in \mathbb{R}$. We say (x_n) has *order of convergence* $p \in \mathbb{R}^+$ if exists $C > 0$ such that:

$$\lim_{n\to\infty} \frac{|x_{n+1} - \alpha|}{|x_n - \alpha|^p} = C.$$

The constant C is called *asymptotic error constant*. For the case $p = 1$, we need $C < 1$. In this case the convergence is called *linear convergence*; for $p = 2$, is called *quadratic convergence*; for $p = 3$, *cubic convergence*... If it's satisfied that

$$\lim_{n\to\infty} \frac{|x_{n+1} - \alpha|}{|x_n - \alpha|^p} = 0$$

for some $p \in \mathbb{R}^+$, we say the sequence has *order of convergence at least* p.

Theorem 1.27. Let $g : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^p$ and let α be a fixed point of g. Suppose

$$g'(\alpha) = g''(\alpha) = \cdots = g^{(p-1)}(\alpha) = 0$$

with $|g'(\alpha)| < 1$ if $p = 1$. Then the iteration $x_{n+1} = g(x_n)$, with x_0 sufficiently close to α, has order of convergence at least p. If, moreover, $g^{(p)}(\alpha) \neq 0$, then the previous iteration has order of convergence p with asymptotic error constant $C = \frac{|g^{(p)}(\alpha)|}{p!}$.

Theorem 1.28. Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^3$ and α be a simple zero of f. If $f''(\alpha) \neq 0$, then Newton-Raphson method for finding α has quadratic convergence with asymptotic error constant $C = \frac{1}{2}\left|\frac{f''(\alpha)}{f'(\alpha)}\right|$.
If $f \in \mathcal{C}^{m+2}$, and α is a zero of multiplicity $m > 1$, then Newton-Raphson method has linear convergence but Newton-Raphson modified method has at least quadratic convergence.

Theorem 1.29. Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^3$ and let α be a simple zero of f. Then Chebyshev's method for finding α has at least cubic convergence.

Definition 1.30. We define the *computational efficiency of an algorithm* as a function $E(p, t)$, where t is the time taken for each iteration of the method and p is the order of convergence of the method. $E(p, t)$ must satisfy the following properties:

1. $E(p, t)$ is increasing with respect to the variable p and decreasing with respect to t.
2. $E(p, t) = E(p^m, mt)$ $\forall m \in \mathbb{R}$.

Examples of such functions are the following:

$$E(p, t) = \frac{\log p}{t}, \quad E(p, t) = p^{1/t}.$$

Sequence acceleration

Definition 1.31 (Aitken's Δ^2 method). Let (x_n) be a sequence of real numbers. We denote:

$$\Delta x_n := x_{n+1} - x_n,$$
$$\Delta^2 x_n := \Delta x_{n+1} - \Delta x_n = x_{n+2} - 2x_{n+1} + x_n.$$

Aitken's Δ^2 method is the transformation of the sequence (x_n) into a sequence y_n, defined as:

$$y_n := x_n - \frac{(\Delta x_n)^2}{\Delta^2 x_n} = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n},$$

with $y_0 = x_0$.

Theorem 1.32. Let (x_n) be a sequence of real numbers such that $\lim_{n\to\infty} x_n = \alpha$, $x_n \neq \alpha$ $\forall n \in \mathbb{N}$ and $\exists C$, $|C| < 1$, satisfying

$$x_{n+1} - \alpha = (C + \delta_n)(x_n - \alpha), \quad \text{with } \lim_{n\to\infty} \delta_n = 0.$$

Then the sequence (y_n) obtained from Aitken's Δ^2 process is well-defined and

$$\lim_{n\to\infty} \frac{y_n - \alpha}{x_n - \alpha} = 0^{5}.$$

[5]This means that Aitken's Δ^2 method produces an acceleration of the convergence of the sequence (x_n).

Theorem 1.33 (Steffensen's method). Let $g:\mathbb{R}\to\mathbb{R}$ be a continuous function and suppose we have an iterative method $x_{n+1}=g(x_n)$. Then for each step n we can consider a new iteration y_{n+1}, with $y_0=x_0$, given by:

$$y_{n+1}=y_n-\frac{(g(y_n)-y_n)^2}{g(g(y_n))-2g(y_n)+y_n}.$$

Proposition 1.34. Let $f:\mathbb{R}\to\mathbb{R}$ be a function of class $\mathcal{C}^2$ and α be a simple zero of f. Then Steffensen's method for finding α has at least quadratic convergence[6].

Zeros of polynomials

Lemma 1.35. Let $p(z)=a_0+a_1z+\cdots+a_nz^n\in\mathbb{C}[x]$ with $a_n\neq 0$. We define

$$\lambda:=\max\left\{\left\|\frac{a_i}{a_n}\right\|:i=0,1,\ldots,n-1\right\}.$$

Then if $p(\alpha)=0$ for some $\alpha\in\mathbb{C}$, $\|\alpha\|\leq\lambda+1$.

Definition 1.36 (Strum's sequence). Let (f_i), $i=0,\ldots,n$, be a sequence of continuous functions defined on $[a,b]\subset\mathbb{R}$ and $f:[a,b]\to\mathbb{R}$ be a function of class $\mathcal{C}^1$ such that $f(a)f(b)\neq 0$. We say (f_n) is a *Sturm's sequence* if:

1. $f_0=f$.
2. If $\alpha\in[a,b]$ satisfies $f_0(\alpha)=0\implies f_0'(\alpha)f_1(\alpha)>0$.
3. For $i=1,\ldots,n-1$, if $\alpha\in[a,b]$ satisfies $f_i(\alpha)=0\implies f_{i-1}(\alpha)f_{i+1}(\alpha)<0$.
4. $f_n(x)\neq 0\ \forall x\in[a,b]$.

Definition 1.37. Let (a_i), $i=0,\ldots,n$, be a sequence. We define $\nu(a_i)$ as the number of sign variations of the sequence

$$\{a_0,a_1,\ldots,a_n\},$$

without taking into account null values.

Theorem 1.38 (Sturm's theorem). Let $f:[a,b]\to\mathbb{R}$ be a function of class $\mathcal{C}^1$ such that $f(a)f(b)\neq 0$ and with a finite number of zeros. Let (f_i), $i=0,\ldots,n$, be a Sturm sequence defined on $[a,b]$. Then the number of zeros of f on $[a,b]$ is

$$\nu\left(f_i(a)\right)-\nu\left(f_i(b)\right).$$

Lemma 1.39. Let $p\in\mathbb{C}[x]$ be a polynomial. Then the polynomial $q=\dfrac{p}{\gcd(p,p')}$ has the same roots as p but all of them are simple.

Proposition 1.40. Let $p\in\mathbb{R}[x]$ be a polynomial with $\deg p=m$. We define $f_0=\dfrac{p}{\gcd(p,p')}$ and $f_1=f_0'$. If $\deg f_0=n$, then for $i=0,1,\ldots,n-2$, we define f_{i+2} as

$$f_i(x)=q_{i+1}(x)f_{i+1}(x)-f_{i+2}(x),$$

(similarly to the euclidean division between f_i and f_{i+1}). Then f_n is constant and hence the sequence (f_i), $i=0,\ldots,n$, is a Sturm sequence.

Theorem 1.41 (Budan-Fourier theorem). Let $p\in\mathbb{R}[x]$ be a polynomial with $\deg p=n$. Consider the sequence $(p^{(i)})$, $i=0,\ldots,n$. If $p(a)p(b)\neq 0$, the number of zeros of p on $[a,b]$ is

$$\nu\left(p^{(i)}(a)\right)-\nu\left(p^{(i)}(b)\right)-2k,\quad\text{for some }k\in\mathbb{N}\cup\{0\}.$$

Corollary 1.42 (Descartes' rule of signs). Let $p=a_0+a_1x+\cdots+a_nx^n\in\mathbb{R}[x]$ be a polynomial. If $p(0)\neq 0$, the number of zeros of p on $[0,\infty)$ is

$$\nu(a_i)-2k,\quad\text{for some }k\in\mathbb{N}\cup\{0\}\text{[7]}.$$

Theorem 1.43 (Greshgorin theorem). Let $A=(a_{ij})\in\mathcal{M}_n(\mathbb{C})$ be a complex matrix and λ be an eigenvalue of A. For all $i,j\in\{1,2,\ldots,n\}$ we define:

$$r_i=\sum_{\substack{k=1\\k\neq i}}^n|a_{ik}|,\quad R_i=\{z\in\mathbb{C}:|z-a_{ii}|\leq r_i\},$$

$$c_j=\sum_{\substack{k=1\\k\neq j}}^n|a_{kj}|,\quad C_j=\{z\in\mathbb{C}:|z-a_{jj}|\leq c_j\}.$$

Then $\lambda\in\bigcup_{i=1}^n R_i$ and $\lambda\in\bigcup_{j=1}^n C_j$. Moreover in each connected component of $\bigcup_{i=1}^n R_i$ (respectively $\bigcup_{j=1}^n C_j$) there are as many eigenvalues (taking into account the multiplicity) as disks R_i (respectively C_i).

Corollary 1.44. Let $p(z)=a_0+a_1z+\cdots+a_nz^n+z^{n+1}\in\mathbb{C}[x]$. We define

$$r=\sum_{i=1}^{n-1}|a_i|,\quad c=\max\{|a_0|,|a_1|+1,\ldots,|a_{n-1}|+1\}.$$

Then if $p(\alpha)=0$ for some $\alpha\in\mathbb{C}$,

$$\alpha\in(B(0,1)\cup B(-a_n,r))\cap(B(-a_n,1)\cup B(0,c)).$$

0.1.3 | Interpolation

Definition 1.45. Suppose we have a family of real valued functions $\mathfrak{C}$ and a set of points $\{(x_i,y_i)\}_{i=0}^n:=\{(x_i,y_i):x_j\neq x_k\iff j\neq k,i=0,\ldots,n\}$. These points $\{(x_i,y_i)\}_{i=0}^n$ are called *support points*. The *interpolation problem* consists in finding a function $f\in\mathfrak{C}$ such that $f(x_i)=y_i$ for $i=0,\ldots,n$[8].

[6] Note that the advantage of Steffensen's method over Newton-Raphson method is that in the former we don't need the differentiability of the function whereas in the latter we do.

[7] Note that making the change of variable $t=-x$ one can obtain the number of zeros on $(-\infty,0]$ of p by considering the polynomial $p(t)$.

[8] Types of interpolation are for example polynomial interpolation, trigonometric interpolation, Padé interpolation, Hermite interpolation and spline interpolation.

Polynomial interpolation

Definition 1.46. Given a set of support points $\{(x_i, y_i)\}_{i=0}^n$, *Lagrange's interpolation problem* consists in finding a polynomial $p_n \in \mathbb{R}[x]$ such that $\deg p_n \leq n$ and $p_n(x_i) = y_i$.

Proposition 1.47. Lagrange's interpolation problem has a unique solution and this is:

$$p_n(x) = \sum_{k=0}^{n} y_k \frac{\omega_n(x)}{\omega_n'(x)}, \quad \text{where } \omega_n(x) := \prod_{j=0}^{n}(x - x_j).$$

Proposition 1.48 (Neville's algorithm). Let $P_{i_1,\ldots,i_k}(x) \in \mathbb{R}[x]$ be such that $\deg P_{i_0,\ldots,i_k} \leq k$ and $P_{i_1,\ldots,i_k}(x_{i_j}) = y_{i_j}$ for $j = 0, \ldots, k$. Then, it is satisfied that:

1. $P_i(x) = y_i$.
2. $P_{i_0,\ldots,i_k}(x) = \dfrac{\begin{vmatrix} P_{i_1,\ldots,i_k}(x) & x - x_{i_k} \\ P_{i_0,\ldots,i_{k-1}}(x) & x - x_{i_0} \end{vmatrix}}{x_{i_k} - x_{i_0}}$

Definition 1.49. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $\{x_i\}_{i=0}^k \subset \mathbb{R}$ be pairwise distinct points. We define the *divided difference of order k of f applied to* $\{x_i\}_{i=0}^k$, denoted by $f[x_0, \ldots, x_k]$, as the coefficient of x^k of the interpolating polynomial with support points $\{(x_i, f(x_i))\}_{i=0}^k$

Proposition 1.50. Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $\{x_i\}_{i=0}^k \subset \mathbb{R}$ be different points. Lagrange interpolating polynomial with support points $\{(x_i, f(x_i))\}_{i=0}^k$ is

$$p_n(x) = \sum_{j=0}^{n} f[x_j]\omega_{j-1}(x),$$

assuming $\omega_{-1} := 1$.

Proposition 1.51 (Newton's divided differences method). Let $f : \mathbb{R} \to \mathbb{R}$ be a function. For $x \in \mathbb{R}$, we have $f[x] = f(x)$. And if $\{x_i\}_{i=0}^n \subset \mathbb{R}$ are different points, then

$$f[x_0, \ldots, x_n] = \frac{f[x_1, \ldots, x_n] - f[x_0, \ldots, x_{n-1}]}{x_n - x_0}.$$

Theorem 1.52. Let $f : [a, b] \to \mathbb{R}$ be a function of class $\mathcal{C}^{n+1}$, $\{x_i\}_{i=0}^n \subset \mathbb{R}$ be different points and $p_n \in \mathbb{R}[x]$ be the interpolating polynomial with support points $\{(x_i, f(x_i))\}_{i=0}^n$. Then $\forall x \in [a, b]$:

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}\omega_n(x),$$

where $\xi_x \in \langle x_0, \ldots, x_n, x\rangle$[9].

Lemma 1.53. Let $f : [a, b] \to \mathbb{R}$ be a function of class $\mathcal{C}^{n+1}$ and $\{x_i\}_{i=0}^n \subset \mathbb{R}$ be pairwise distinct points. Then $\exists \xi \in \langle x_0, \ldots, x_n\rangle$ such that:

$$f[x_0, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!}.$$

Proposition 1.54. Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^{n+1}$, $\{x_i\}_{i=0}^n \subset \mathbb{R}$ be pairwise distinct points and $\sigma \in S_n$. Then

$$f[x_0, \ldots, x_n] = f[x_{\sigma(0)}, \ldots, x_{\sigma(n)}]$$

Definition 1.55. Let $\{(x_i, y_i)\}_{i=0}^n$ be support points. The x-axis points $\{(x_i)\}_{i=0}^n$ are *equally-spaced* if

$$x_i = x_0 + ih, \quad \text{for } i = 0, \ldots, n \quad \text{with} \quad h := \frac{x_n - x_0}{n}.$$

Definition 1.56. Let $f : [a, b] \to \mathbb{R}$ be a function and $\{x_i\}_{i=0}^n \subset \mathbb{R}$ be equally-spaced points. We define:

$$\Delta f(x) := f(x + h) - f(x),$$
$$\Delta^{n+1} f(x) := \Delta(\Delta^n f(x)).$$

Lemma 1.57. Let $f : [a, b] \to \mathbb{R}$ be a function and $\{x_i\}_{i=0}^n \subset \mathbb{R}$ be equally-spaced points. Then,

$$\Delta^n f(x_0) = n! h^n f[x_0, \ldots, x_n].$$

Corollary 1.58. Let $f \in \mathbb{R}[x]$ with $\deg f = n$. Suppose we interpolate f with equally-spaced nodes. Then, $\Delta^n f(x) \equiv$ constant.

Hermite interpolation

Definition 1.59. Given a sets of points $\{(x_i)\}_{i=0}^m \subset \mathbb{R}$, $\{(n_i)\}_{i=0}^m \subset \mathbb{N}$ and $\{(y_i^{(k)} : k = 0, \ldots, n_i - 1\}_{i=0}^m \subset \mathbb{R}$ *Hermite interpolation problem* consists in finding a polynomial $h_n \in \mathbb{R}[x]$ such that $\deg h_n \leq n$, $\sum_{i=0}^m n_i = n + 1$ and

$$h_n^{(k)}(x_i) = y_i^{(k)} \text{ for } i = 0, \ldots, m \text{ and } k = 0, \ldots, n_i - 1.$$

Proposition 1.60. Hermite interpolation problem has a unique solution.

Theorem 1.61. Let $f : [a, b] \to \mathbb{R}$ be a function of class $\mathcal{C}^{n+1}$, $\{x_i\}_{i=0}^m \subset \mathbb{R}$ be different points, $\{(n_i)\}_{i=0}^m \subset \mathbb{N}$ be such that $\sum_{i=0}^m n_i = n + 1$. Let h_n be the Hermite interpolating polynomial of f with nodes $\{x_i\}_{i=0}^m \subset \mathbb{R}$, that is,

$$h_n^{(k)}(x_i) = f^{(k)}(x_i) \text{ for } i = 0, \ldots, m \text{ and } k = 0, \ldots, n_i - 1.$$

Then $\forall x \in [a, b]$ $\exists \xi_x \in \langle x_0, \ldots, x_n, x\rangle$ such that:

$$f(x) - h_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}(x - x_0)^{n_0} \cdots (x - x_m)^{n_m}.$$

[9]The interval $\langle a_1, \ldots, a_k\rangle$ is defined as $\langle a_1, \ldots, a_k\rangle := (\min(a_1, \ldots, a_k), \max(a_1, \ldots, a_k))$.

Spline interpolation

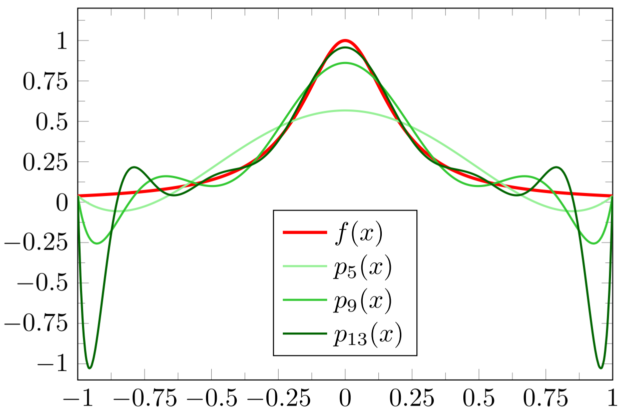

Figure 2: Runge's phenomenon. In this case $f(x) = \frac{1}{1+25x^2}$. $p_5(x)$ is the 5th-order Lagrange interpolating polynomial with equally-spaced interpolating points; $p_9(x)$, the 9th-order Lagrange interpolating polynomial with equally-spaced interpolating points, and $p_{13}(x)$, the 13th-order Lagrange interpolating polynomial with equally-spaced interpolating points.

Definition 1.62 (Spline). Let $\{(x_i, y_i)\}_{i=0}^n$ be support points of an interval $[a, b]$. A *spline of degree* p is a function $s : [a, b] \to \mathbb{R}$ of class $\mathcal{C}^{p-1}$ satisfying:

$$s_{|[x_i,x_{i+1}]} \in \mathbb{R}[x], \quad \deg s_{|[x_i,x_{i+1}]} = p, \quad s(x_i) = y_i,$$

for $i = 0, \ldots, n-1$. The most common case are splines of degree $p = 3$ or *cubic spline*. In this case we can impose two more conditions on their definition in one of the following ways:

1. *Natural cubic spline*:
$$s''(x_0) = s''(x_n) = 0.$$
2. *Cubic Hermite spline*: Given $y_0', y_n' \in \mathbb{R}$,
$$s'(x_0) = y_0', \quad s'(x_n) = y_n'.$$
3. *Cubic periodic spline*:
$$s'(x_0) = s'(x_n), \quad s''(x_0) = s''(x_n)$$

Definition 1.63. Let $f : [a, b] \to \mathbb{R}$ a function of class $\mathcal{C}^2$. We define the *seminorm*[10] *of* f as

$$\|f\|^2 = \int_a^b (f''(x))^2 dx.$$

Proposition 1.64. Let $f : [a, b] \to \mathbb{R}$ a function of class $\mathcal{C}^2$ interpolating the support points $\{(x_i, y_i)\}_{i=0}^n \subset \mathbb{R}^2$, $a \leq x_0 < \cdots < x_n \leq b$. If s is the natural cubic spline associated with $\{(x_i, y_i)\}_{i=0}^n$, then:

$$\|f-s\|^2 = \|f\|^2 - \|s\|^2 - 2(f'-s)s''\Big|_{x_0}^{x_n} + 2\sum_{i=1}^n (f-s)s'''\Big|_{x_{i-1}^+}^{x_i^-}.$$

Theorem 1.65. Let $f : [a, b] \to \mathbb{R}$ a function of class $\mathcal{C}^2$ interpolating the support points $\{(x_i, y_i)\}_{i=0}^n \subset \mathbb{R}^2$, $a \leq x_0 < \cdots < x_n \leq b$. If s is the natural cubic spline associated with $\{(x_i, y_i)\}_{i=0}^n$, then

$$\|s\| \leq \|f\|.$$

0.1.4 | Numerical differentiation and integration

Differentiation

Theorem 1.66 (Intermediate value theorem). Let $f : [a, b] \to \mathbb{R}$ be a continuous function, $\xi_0, \ldots, \xi_n \in [a, b]$ and $\alpha_0, \ldots, \alpha_n \geq 0$. Then, $\exists \eta \in [a, b]$ such that:

$$\sum_{i=0}^n \alpha_i f(\xi_i) = \left(\sum_{i=0}^n \alpha_i\right) f(\eta).$$

Theorem 1.67 (Forward and backward difference formula of order 1). Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^2$. Then, forward difference formula of order 1 is:

$$f'(a) = \frac{f(a+h) - f(a)}{h} - \frac{f''(\xi)}{2}h,$$

where $\xi \in \langle a, a+h \rangle$. Analogously, backward difference formula of order 1 is:

$$f'(a) = \frac{f(a) - f(a-h)}{h} + \frac{f''(\eta)}{2}h,$$

where $\eta \in \langle a-h, a \rangle$.

Theorem 1.68 (Symmetric difference formula of order 1). Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^3$. Then, symmetric difference formula of order 1:

$$f'(a) = \frac{f(a+h) - f(a-h)}{2h} - \frac{f^{(3)}(\xi)}{6}h^2,$$

where $\xi \in \langle a-h, a+h \rangle$.

Theorem 1.69 (Symmetric difference formula of order 2). Let $f : \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^4$. Then, symmetric difference formula of order 2:

$$f''(a) = \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} - \frac{f^{(4)}(\xi)}{12}h^2,$$

where $\xi \in \langle a-h, a, a+h \rangle$.

Richardson extrapolation

Theorem 1.70 (Richardson extrapolation). Suppose we have a function f that approximate a value α with an error that depends on a small quantity h. That is:

$$f(h) = \alpha + a_1 h^{k_1} + a_2 h^{k_2} + \cdots,$$

with $k_1 < k_2 < \cdots$ and a_i are unknown constants. Given $q > 0$, we define

$$D_1(h) = f(h), \quad D_{n+1}(h) = \frac{q^{k_n} D_n(h/q) - D_n(h)}{q^{k_n} - 1}.$$

And we can observe that $\alpha = D_{n+1}(h) + O(h^{k_{n+1}})$.

[10] The term *seminorm* has been used instead of *norm* to emphasize that not all properties of a norm are satisfied with this definition.

Integration

Definition 1.71. Let $f : [a, b] \to \mathbb{R}$ be a continuous function, $\{x_i\}_{i=0}^n \subseteq [a, b]$ be a set of nodes and P_n be the Lagrange interpolating polynomial with support points $\{(x_i, f(x_i))\}_{i=0}^n$. We define the *integration formula base on interpolation* as

$$I(f) = \int_a^b P_n(x)dx \tag{1}$$

Lemma 1.72. Let $f : [a, b] \to \mathbb{R}$ be a continuous function $\{x_i\}_{i=0}^n \subseteq [a, b]$ be a set of nodes. Then,

$$I(f) = \sum_{i=1}^{n} A_i f(x_i), \quad \text{where } A_i = \int_a^b \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} dx.$$

Lemma 1.73. Let $p \in \mathbb{R}[x]$ be a polynomial defined on an interval $[a, b]$ such that $\deg p \leq n$ and let $\{x_i\}_{i=0}^n \subseteq [a, b]$ be a set of nodes. Then, $I(p) = \int_a^b p(x)dx$.

Lemma 1.74. Let $p \in \mathbb{R}[x]$ be a polynomial defined on an interval $[a, b]$ such that $\deg p \leq n$ and let $\{x_i\}_{i=0}^n \subseteq [a, b]$ be a set of nodes. Then,

$$I(p) = \int_a^b p(x)dx \iff I(x^k) = \int_a^b x^k dx, \text{ for } 0 \leq k \leq n.$$

Newton-Cotes formulas