

## APPROXIMATION TECHNIQUES

### 1. Rate and order of convergence

The objective of scientific computation is to solve some problem  $f(x) = 0$  by constructing a sequence of approximations  $\{x_n\}_{n \in \mathbb{N}}$ . The condition suggested by mathematical analysis would be  $x = \lim_{n \rightarrow \infty} x_n$ , with  $f(x) = 0$ . As already in the Leibniz series approximation of  $\pi$ , acceptable accuracy might only be obtained for large  $n$ . Since  $f$  could be an arbitrarily complex mathematical object, such slowly converging approximating sequences are of little practical interest. Scientific computing seeks approximations of the solution with rapidly decreasing error. This change of viewpoint with respect to analysis is embodied in the concepts of *rate* and *order of convergence*.

DEFINITION 1.  $\{x_n\}_{n \in \mathbb{N}}$  converges to  $x$  with rate  $r \in (0, 1)$  and order  $p$  if

$$\lim_{n \rightarrow \infty} \frac{|x_{n+1} - x|}{|x_n - x|^p} = r. \quad (1)$$

As previously discussed, the above definition is of limited utility since:

- a) The solution  $x$  is unknown;
- b) The limit  $n \rightarrow \infty$  is impractical to attain.

Sequences converge faster for higher order  $p$ , but lower rate  $r$ . A more useful approach is to determine estimates of the rate and order of convergence over some range of iterations that are sufficiently accurate. Rewriting (1) as

$$\lim_{n \rightarrow \infty} (|x_{n+1} - x| - r|x_n - x|^p) = 0,$$

suggests introducing the distance between successive iterates  $d_n = |x_n - x_{n-1}|$ , and considering the condition

$$|d_{n+1} - s d_n^q| \text{ small for large } n.$$

DEFINITION 2.  $\{x_n\}_{n \in \mathbb{N}}$  approximates  $x$  with rate  $s$  and order  $q$  if there exist  $s, q \in \mathbb{R}$  and  $n_1, n_2 \in \mathbb{N}$  such that

$$|d_{n+1} - s d_n^q| < \epsilon, \text{ for } n_1 \leq n \leq n_2 \quad (2)$$

with  $d_n = |x_n - x_{n-1}|$ ,  $n \in \mathbb{N}$ ,  $\epsilon$  denotes machine epsilon.

As an example, consider the derivative  $g = f'$  of  $f(x) = e^x - 1$  at  $x_0 = 0$ , as given by the calculus definition

$$g(x_0) = f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h},$$

and construct a sequence of approximations

$$g_n = \frac{f_n - f(0)}{h_n}, f_n = f(h_n), h_n = 2^{-n}.$$

Start with a numerical experiment, and compute the sequence  $d_n = |g_n - g_{n-1}|$ .

$n$	1	2	...	$N$
$h_n = 2^{-n}$	1/2	1/4	...	$1/2^N$
$f_n$	$f_1$	$f_2$	...	$f_N$
$g_n = (f_n - f(0)) / h_n$	$g_1 = (f_1 - f(0)) / 2$	$g_2 = (f_2 - f(0)) / 4$	...	$g_N = (f_N - f(0)) / 2^N$
$d_{n-1} =  g_n - g_{n-1} $	-	$d_1 =  g_2 - g_1 $	...	$d_{N-1} =  g_N - g_{N-1} $

**Table 1.** Table presentation of calculations to construct approximation of derivative sequence for  $f(x) = \tan x$ , at  $x_0 = 0$ .

```

∴ N=24; n=1:N; h=2.0.^(-n); f(x) = exp(x)-1; x0=0; f0=f(x0);
∴ g = (f.(h).-f0) ./ h; d=abs.(g[2:N]-g[1:N-1]);
∴ n1=2; n2=8; [h[n1:n2] g[n1:n2] d[n1:n2]]

```

$$\begin{bmatrix}
0.25 & 1.1361016667509656 & 0.070914042216355 \\
0.125 & 1.0651876245346106 & 0.033276281848861444 \\
0.0625 & 1.0319113426857491 & 0.0161223027144608 \\
0.03125 & 1.0157890399712883 & 0.007935690423401809 \\
0.015625 & 1.0078533495478865 & 0.0039369071225365815 \\
0.0078125 & 1.00391644242535 & 0.001960771808398931 \\
0.00390625 & 1.001955670616951 & 0.0009784720235188615
\end{bmatrix} \quad (3)$$

```

∴

```

Investigation of the numerical results indicates increasing accuracy in the estimate of  $g(x) = (e^x - 1)' = e^x$ ,  $g(0) = 1$  with decreasing step size  $h$ . The distance between successive approximation sequence terms  $d_n = |g_n - g_{n-1}|$  also decreases. It is more intuitive to analyze convergence behavior through a plot rather than a numerical table.

```

∴ clf(); plot(h[2:N], d, "-o"); xlabel("h"); ylabel("d");
∴ cd(homedir() * "//courses//MATH661//images"); savefig("L02Fig01a.eps");
∴

```

The intent of the rate and order of approximation definitions is to state that the distance between successive terms behaves as

$$d_{n+1} \cong s d_n^q,$$

in the hope that this is a Cauchy sequence, and successively closer terms actually indicate convergence. The convergence parameters  $(s, q)$  can be isolated by taking logarithms,  $c_n = \log d_n$  leading to a linear dependence

$$c_{n+1} \cong q c_n + \log s.$$

Subtraction of successive terms gives  $c_n - c_{n-1} \cong q(c_{n-1} - c_{n-2})$ , leading to an average slope estimate

$$q \cong \frac{1}{N-3} \sum_{n=3}^{N-1} \frac{c_n - c_{n-1}}{c_{n-1} - c_{n-2}}$$

```

∴ c=log.(2,d); lh=log.(2,h[2:N]); clf(); plot(lh,c,"-o");
plot([-10,-20],[-10,-20],"k"); plot([-10,-20],[-10,-30],"g");
∴ xlabel("log(h)"); ylabel("log(d)"); savefig("L02Fig01b.eps");
∴ num=c[3:N-1]-c[2:N-2]; den=c[2:N-2]-c[1:N-3];
∴ q = sum(num ./ den) / (N-3)

```

0.9920966582673338

```

∴

```

The above computations indicate  $q \cong 1$ , known as *linear convergence*. Figure 1b shows the common practice of depicting guide lines of slope 1 (black) and slope 2 (green) to visually ascertain the rate of convergence. Once the order of approximation  $q$  is determined, the rate of approximation is estimated from

$$\log s \cong \frac{1}{N-2} \sum_{n=2}^{N-1} (c_n - q c_{n-1}).$$

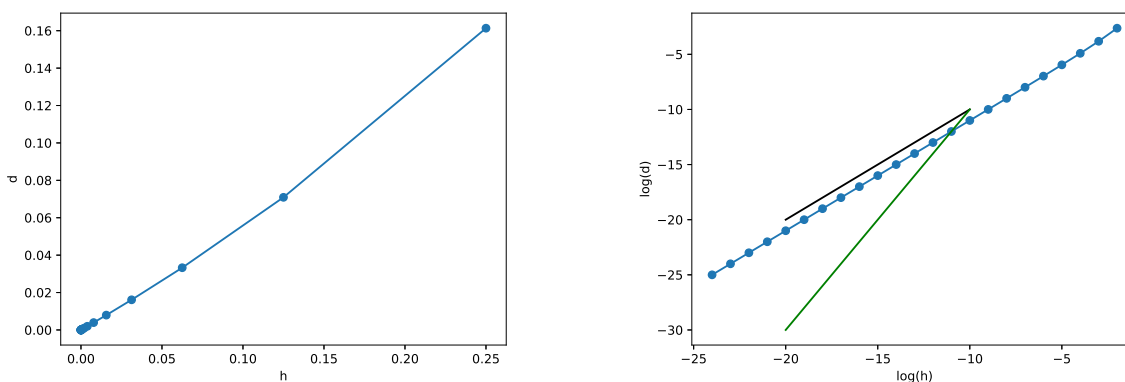
```
∴ s=exp (sum (c [2:N-1] -q*c [1:N-2]) / (N-2))
```

```
0.3252477724180383
```

```
∴
```

The above results suggest successive approximants become closer spaced according to

$$d_n \cong 0.124d_{n-1}$$



**Figure 1.** (a, left). Convergence plot; (b, right) Convergence plot in logarithmic coordinates.

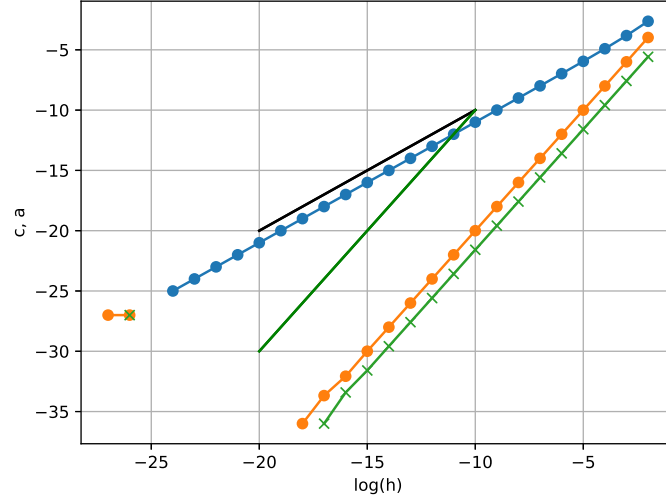
Repeat the above experiment at  $x_0 = \ln 2$ , where  $g(\ln 2) = 2$ , and using a different approximation of the derivative

$$g_n = \frac{f(\ln 2 + h_n) - f(\ln 2 - h_n)}{2h_n}.$$

For this experiment, in addition to the rate and order of approximation ( $s, q$ ), also determine the rate and order of convergence ( $r, p$ ) using

$$b_n = |g_n - g(\pi/4)|, b_{n+1} \cong r b_n^p, a_n = \log b_n, a_{n+1} = p a_n + \log r.$$

```
∴ N=32; n=1:N; h=2.0.^(-n); f(x) = exp(x)-1; x0=log(2); f0=f(x0);
g0=exp(x0);
∴ g = (f.(x0 .+ h) .- f.(x0 .- h)) ./ (2*h); d=abs.(g[2:N]-g[1:N-1]);
∴ c=log.(2,d); lh=log.(2,h[2:N]); b=abs.(g[2:N]-g0); a=log.(2,b);
∴ plot(lh,c,"-o"); plot(lh,a,"-x"); plot([-10,-20],[-10,-20],"k");
plot([-10,-20],[-10,-30],"g");
∴ xlabel("log(h)"); ylabel("c, a"); grid("on");
∴ savefig("L02Fig02.eps");
∴
```



**Figure 2.** Typical convergence behavior for approximants of a derivative. Blue line shows first-order or linear convergence of approximation  $f'(x_0) \cong (f(x_0+h) - f(x_0))/h$  for  $f(x) = e^x - 1$  at  $x_0=0$ . The convergence curve is monotone, with decreasing error for all sample points due to fortuitous  $f(x_0) = 0$ . Green and orange lines indicate that the orders of convergence and approximation are quadratic for  $f'(x_0) \cong (f(x_0+h) - f(x_0-h))/(2h)$  for  $f(x) = e^x - 1$  at  $x_0 = \log 2$ . Now,  $f(x_0) \neq 0$ , and small differences in the numerator are no longer resolved by the floating point system leading to an increase in the error for  $\log(h) < -20$ . The numerical experiment indicates that order of approximation can be used interchangeably with order of convergence, i.e., closer spacing of successive approximations is often an indication of convergence.

## 2. Convergence acceleration

Given some approximation sequence  $\{x_n\}_{n \in \mathbb{N}}$ ,  $x_n \rightarrow x$ , with  $x$  solution of problem  $f(x) = 0$ , it is of interest to construct a more rapidly convergent sequence  $\{y_n\}_{n \in \mathbb{N}}$ ,  $y_n \rightarrow x$ . Knowledge of the order of convergence  $p$  can be used to achieve this purpose by writing

$$x_n - x \cong r(x_{n-1} - x)^p, x_{n-1} - x \cong r(x_{n-2} - x)^p, \quad (4)$$

and taking the ratio to obtain

$$\frac{x_n - x}{x_{n-1} - x} = \left( \frac{x_{n-1} - x}{x_{n-2} - x} \right)^p. \quad (5)$$

For  $p \in \mathbb{N}$ , the above is a polynomial equation of degree  $p$  that can be solved to obtain  $x$ . Since (4) is an approximation, solving (5) gives an approximation of the exact limit.

### 2.1. Aitken acceleration

One of the widely used acceleration techniques was published by Aitken (1926, but had been in use since Medieval times) for  $p = 1$  in which case (5) gives

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

The above suggests that starting from  $\{x_n\}_{n \in \mathbb{N}}$ , the sequence  $\{a_n\}_{n \in \mathbb{N}}$  with

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

might converge faster towards the limit. Investigate by revisiting the numerical experiment on approximation of the derivative  $g = f'$  of  $f(x) = e^x - 1$  at  $x_0 = 0$ , using

$$g_n = \frac{f_n - f(0)}{h_n}, f_n = f(h_n), h_n = 2^{-n}.$$

```

∴ N=24; n=1:N; h=2.0.^(-n); f(x) = exp(x)-1; x0=0; f0=f(x0);
∴ g = (f.(h).-f0) ./ h; a = copy(g);
∴ a[3:N] = g[3:N] - (g[3:N]-g[2:N-1]).^2 ./ (g[3:N]-2*g[2:N-1]+g[1:N-2]);
∴ lh=log.(2,h); d=log.(2,abs.(g.-1)); b=log.(2,abs.(a.-1));
∴ clf(); plot(lh,d,"-o"); plot(lh,b,"-x"); plot([-10,-20],[ -10,-20],"k");
plot([-10,-20],[ -10,-30],"g"); xlabel("log(h)"); ylabel("g, a");
grid("on"); savefig("L02Fig03.eps");
∴

```

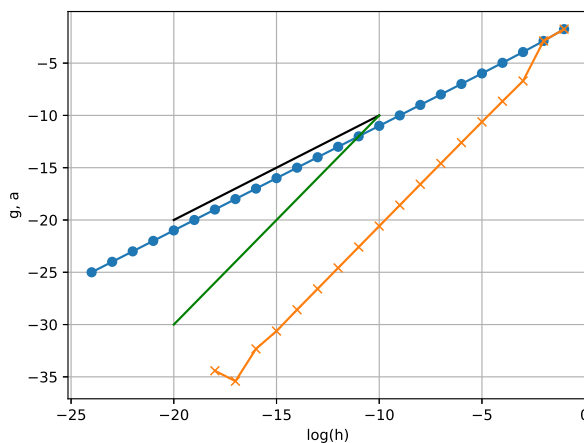


Figure 3. Aitken acceleration of linearly convergent sequence (blue dots) yields a close-to-quadratic convergent sequence (orange x).

### 3. Approximation correction types

Several approaches may be used in construction of an approximating sequence  $\{x_n\}_{n \in \mathbb{N}}$ . The approaches exemplified below for  $x_n \in \mathbb{R}$ , can be generalized when  $x_n$  is some other type of mathematical object.

#### 3.1. Additive corrections

Returning to the Leibniz series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots,$$

the sequence of approximations is  $\{L_n\}_{n \in \mathbb{N}}$  with general term

$$L_n = \sum_{k=0}^n \frac{(-1)^k}{2k+1}.$$

Note that successive terms are obtained by an additive correction

$$L_n = L_{n-1} + \frac{(-1)^n}{2n+1}, L_n \rightarrow \frac{\pi}{4}.$$

Another example, again giving an approximation of  $\pi$  is the Srinivasa Ramanujan series

$$R_n = \frac{2\sqrt{2}}{9801} \sum_{k=0}^n \frac{(4k)! (1103 + 26390k)}{(k!)^4 396^{4k}}, \lim_{n \rightarrow \infty} R_n = \frac{1}{\pi},$$

that can be used to obtain many digits of accuracy with just a few terms.

An example of the generalization of this approach is the Taylor series of a function. For example, the familiar sine power series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots,$$

is analogous, but with rationals now replaced by monomials, and the limit is now a function  $\sin: \mathbb{R} \rightarrow [-1, 1]$ . The general term is

$$T_n(x) = \sum_{k=0}^n \frac{(-1)^k x^{2k+1}}{(2k+1)!},$$

and the same type of additive correction appears, this time for functions,

$$T_n(x) = T_{n-1}(x) + \frac{(-1)^n x^{2n+1}}{(2n+1)!}, T_n(x) \rightarrow \sin x.$$

### 3.2. Multiplicative corrections

Approximating sequences need not be constructed by adding a correction. Consider the approximation of  $\pi/2$  given by Wallis's product (1656)

$$S_n = \left(\frac{2}{1} \cdot \frac{2}{3}\right) \cdot \left(\frac{4}{3} \cdot \frac{4}{5}\right) \cdot \left(\frac{6}{5} \cdot \frac{6}{7}\right) \dots, S_n = \prod_{k=1}^n \frac{4k^2}{4k^2-1}, S_n \rightarrow \frac{\pi}{2},$$

for which

$$S_n = S_{n-1} \cdot \left(\frac{4n^2}{4n^2-1}\right).$$

Another famous example is the Viète formula from 1593

$$\frac{2}{\pi} = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2+\sqrt{2}}}{2} \cdot \frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2} \dots, V_n = \prod_{k=1}^n \frac{\mathbb{N}_{j=1}^k \sqrt{2}}{2}$$

in which the correction is multiplicative with numerators given by nested radicals. Similar to the  $\sum$  symbol for addition, and the  $\prod$  symbol for multiplication, the  $\mathbb{N}$  symbol is used to denote nested radicals

$$\mathbb{N}_{j=1}^k \sqrt{a_j} = \sqrt{a_1 + \sqrt{a_2 + \sqrt{a_3 + \dots + \sqrt{a_k}}}}.$$

In the case of the Viète formula,  $a_j = 2, b_j = 2$  for all  $j$ .

### 3.3. Continued fractions

Yet another alternative is that of continued fractions, with one possible approximation of  $\pi$  given by

$$\pi + 3 = 6 + \frac{1^2}{6 + \frac{3^2}{6 + \frac{5^2}{6 + \dots}}}. \tag{6}$$

A notation is introduced for continued fractions using the  $\mathbb{K}$  symbol

$$F_n = b_0 + \mathbb{K}_{k=1}^n \frac{a_k}{b_k} = b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{b_3 + \dots}}}.$$

Using this notation, the sequences arising in the continued fraction representation of  $\pi$  are  $\{a_n\}_{n \in \mathbb{N}}, \{b_n\}_{n \in \mathbb{N}}$  chosen as  $a_k = (2k-1)^2$  for  $k \in \mathbb{N}_+$ , and  $b_k = 6$  for  $k \in \mathbb{N}$ .

$$\pi = \lim_{n \rightarrow \infty} \left(6 + \mathbb{K}_{k=1}^n \frac{(2k-1)^2}{6}\right).$$

### 3.4. Composite corrections

The above correction techniques used arithmetic operations. The repeated radical coefficients in the Viète formula suggest consideration of repeated composition of arbitrary functions  $t_0, t_1, \dots, t_n$  to construct the approximant

$$T_n = t_0 \circ t_1 \circ \dots \circ t_n = \bigcirc_{k=0}^n t_k.$$

This is now a general framework, in which all of the preceding correction approaches can be expressed. For example, the continued fraction formula (6) is recovered through the functions

$$t_0(z) = 6 + z, t_1(z) = \frac{1}{6 + z}, \dots, t_k(z) = \frac{(2k-1)^2}{6 + z},$$

and evaluation of the composite function at  $z=0$

$$F_n = T_n(0).$$

This general framework is of significant current interest since such composition of nonlinear functions is the basis of deep neural network approximations.

#### Summary.

- The cornerstone of scientific computing is construction of approximating sequences.
- The problem of number approximation leads to definition of concepts and techniques that can be extended to more complex mathematical objects.
- A primary objective is the construction of efficient approximating sequences, with efficiency characterized through concepts such as order and speed of convergence.
- Though often enforced analytically, limiting behavior of the sequence is of secondary interest. As seen in the approximation of a derivative, the approximating sequence might diverge, yet give satisfactory answers for some range of indices.
- Though by far the most widely studied and used approach to approximation, additive corrections are not the only possibility.
- Alternative correction techniques include: multiplication, continued fractions, or repeated function composition.
- Repeated composition of functions is used in constructing deep neural network approximants.