Lecture Notes for Math 5630/6630 **Fall 2024**

Note 10: Extrapolation

Tags: `math.na` *Date: 10/15/2024*

Disclaimer: *This lecture note is for math 5630/6630 class only.*

A vast number of applications such as the calculation of tangent vectors or areas lead to the problem of computing

$$\mathcal{D}(f) := \frac{d}{dx} f(x), \quad \mathcal{I}(f) := \int_a^b f(x)dx,$$

for certain function $f(x) \in C^k([a, b])$. Accurate evaluations would sometimes be difficult if an analytic expression is absent. Especially when the function values of f are only accessible at a finite number of nodes. Therefore, finding simple yet effective methods to approximate the derivatives and integrals is important.

1 Richardson Extrapolation

From the previous discussion, we already know that interpolation provides an estimate *within* the original observation range. The extrapolation is similar but aims to produce estimates outside the observation range. However, extrapolation may be subject to a greater uncertainty (Fig 10.1), one should use it only when an overestimate is hardly occurring. Suppose there is a sequence

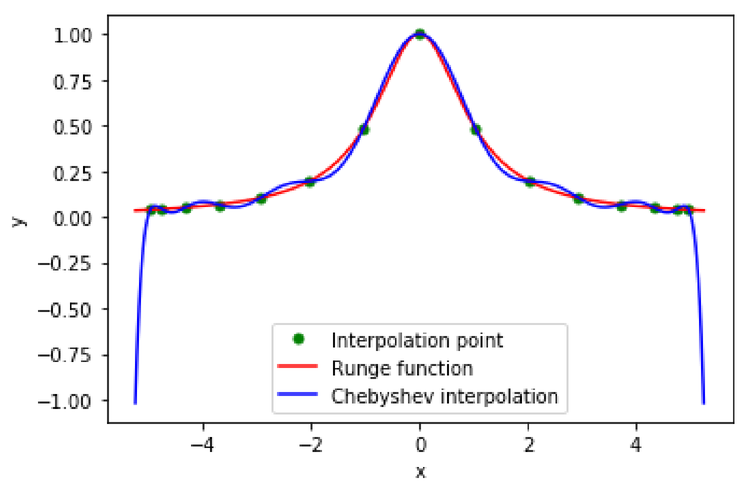

Figure 10.1: Extrapolation behavior for Chebyshev interpolation with 15 nodes.

of estimates $A(h)$ depending on the parameter h smoothly, the limit $A^* = \lim_{h\to 0^+} A(h)$ is the quantity to be computed. In practice, we only have access to $A(h)$ for a few values of h. Using these values to estimate A^* is a typical problem in extrapolation.

The basic idea behind *Richardson extrapolation* is to use polynomial interpolation with a sequence of nodes $h_j \to 0$. Suppose that the function $A(h)$ admits the following asymptotic expansion:

$$A(h) = a_0 + a_1 h^\gamma + a_2 h^{2\gamma} + \cdots + a_k h^{k\gamma} + \mathcal{O}(h^{(k+1)\gamma})$$

for any $h > 0$ and $k \geq 0$. Then $A^* = a_0$ and $A(h) = A^* + \mathcal{O}(h^\gamma)$. Suppose we have access to the values $A(h_0), \ldots, A(h_n)$, then this uniquely determines a polynomial $f_n \in \Pi_n$ and $f_n(h_j^\gamma) = A(h_j)$. We will approximate $A(0) \approx f_n(0)$. The computation of f_n follows the construction of the Newton form.

Lemma 1.1. *Suppose h_j can be represented as*

$$h_j = \frac{\hbar}{t_j}$$

for some adjustable parameter $\hbar$ and scaling constants $1 < t_0 < t_1 < \cdots < t_{n-1}$. Then

$$f_n(0) = A^* + (-1)^n \frac{a_{n+1}}{\prod_{j=0}^n t_j^\gamma} \hbar^{(n+1)\gamma} + \mathcal{O}(\hbar^{(n+2)\gamma}), \quad \text{as } \hbar \to 0.$$

Proof. We view $A(h)$ as a polynomial with respect to h^γ of degree $(n+1)$ with an addition perturbation $\mathcal{O}(h^{(n+2)\gamma})$. Then we have the following.

$$A(h) = p_{n+1}(h^\gamma) + \mathcal{O}(h^{(n+2)\gamma}).$$

Let $\tilde{f}_n$ be the interpolation polynomial of degree n to p_{n+1},

$$p_{n+1}(x) \equiv \tilde{f}_n(x) + p[x, x_0, x_1, \ldots, x_n] \prod_{j=0} (x - h_j^\gamma).$$

where $p[x, x_0, x_1, \ldots, x_n]$ is the coefficient of the leading power in p_{n+1}, a_{n+1}. Thus,

$$A^* = p_{n+1}(0) = \tilde{f}_n(0) + a_{n+1} \prod_{j=0} (0 - h_j^\gamma).$$

Use the result we have discussed in the stability of polynomial interpolation. Therefore,

$$|\tilde{f}_n(0) - f_n(0)| \leq \lambda_n(0) \cdot \mathcal{O}(\hbar^{(n+2)\gamma})$$

Here the Lebesgue function at zero $\lambda_n(0)$ is

$$\lambda_n(0) = \sum_{j=0}^n \prod_{k=0, k\neq j}^n \left| \frac{h_k^\gamma}{h_k^\gamma - h_j^\gamma} \right| = \sum_{j=0}^n \prod_{k=0, k\neq j}^n \left| \frac{1}{1 - (\frac{t_k}{t_j})^\gamma} \right|,$$

which is independent of $\hbar$. $\square$

The Richardson extrapolation considers the special choice of $t_j = t^j$ for some $t > 1$. The error estimate then is

$$f_n(0) = A^* + \left(\frac{(-1)^n}{t^{n(n+1)\gamma/2}} a_{n+1} \right) \hbar^{(n+1)\gamma} + \mathcal{O}(\hbar^{(n+2)\gamma}).$$

There are easier ways to calculate the Richardson extrapolation using the following expansion.

$$A(h) - t^\gamma A\left(\frac{h}{t}\right) = (1 - t^\gamma)A^* + \cancel{a_1 \left(h^\gamma - t^\gamma \left(\frac{h}{t}\right)^\gamma \right)} + a_2 \left(h^{2\gamma} - t^\gamma \left(\frac{h}{t}\right)^{2\gamma} \right) + \dots$$

Let $A_1(h) = \frac{A(h) - t^\gamma A(\frac{h}{t})}{1 - t^\gamma}$, we obtain the first iteration result as

$$A^* \approx A_1(h) + \mathcal{O}(h^{2\gamma}),$$

then follow the same idea, we cancel the $\mathcal{O}(h^{2\gamma})$ term by

$$A_1(h) - t^{2\gamma} A_1\left(\frac{h}{t^2}\right) = (1 - t^{2\gamma})A^* + \mathcal{O}(h^{3\gamma}).$$

Therefore by taking $A_2(h) = \frac{A_1(h) - t^{2\gamma} A_1(\frac{h}{t^{2\gamma}})}{1 - t^{2\gamma}}$, the second iteration satisfies

$$A^* \approx A_2(h) + \mathcal{O}(h^{3\gamma}).$$

However, such a process can constantly refine the approximation due to the potentially fast-growing constant in the $\mathcal{O}$ notation.

2 Aitken Extrapolation

Alexander Aitken rediscovered Aitken extrapolation, which has been used to accelerate a sequence's convergence.

Given a sequence $S = \{s_n\}_{n \ge 0}$, the Aitken extrapolation generates a new sequence

$$AS = \left\{ \frac{s_n s_{n+2} - s_{n+1}^2}{s_n + s_{n+2} - 2s_{n+1}} \right\}_{n \ge 0}.$$

A more stable formulation (why?) is written as

$$(AS)_n = s_n - \frac{(\Delta s_n)^2}{\Delta^2 s_n},$$

where Δ is the forward difference operator that $\Delta s_n = s_{n+1} - s_n$. It is not difficult to see that Aitken extrapolation can accelerate linearly convergent sequences.

Theorem 2.1. *Assume that the sequence $S = \{s_n\}_{n \ge 0}$ satisfies*

$$\lim_{n \to \infty} \frac{|s_{n+1} - \mu|}{|s_n - \mu|} = \rho \in (0, 1).$$

Then the accelerated sequence AS converges faster than S.

Proof. Let $\rho_n := \frac{s_{n+1}-\mu}{s_n-\mu}$, without loss of generality, we assume $\lim_{n\to\infty}\rho_n = \rho$. By the definition of the acceleration formula, we obtain

$$\frac{(AS)_n-\mu}{s_n-\mu} = 1 - \frac{(\Delta s_n)^2}{\Delta^2 s_n}\frac{1}{s_n-\mu}.$$

For sufficiently large n, we have $\rho_n \approx \rho$ and

$$\Delta s_n = s_{n+1}-\mu-(s_n-\mu) = (\rho_n - 1)(s_n-\mu)$$
$$\Delta^2 s_n = s_{n+2}-\mu+s_n-\mu-2(s_{n+1}-\mu) = (\rho_{n+1}\rho_n - 2\rho_n + 1)(s_n-\mu)$$

Therefore

$$\frac{(AS)_n-\mu}{s_n-\mu} = 1 - \frac{(\rho_n-1)^2}{(\rho_{n+1}\rho_n-2\rho_n+1)} = \frac{\rho_n(\rho_{n+1}-\rho_n)}{(\rho_{n+1}\rho_n-2\rho_n+1)} \to 0.$$

□

Remark 2.2. *Let $\varepsilon_n := s_n - \mu$. If the error satisfies the following relation (common in most fixed-point iterations)*

$$\varepsilon_{n+1} = \varepsilon_n(\rho + c_1\varepsilon_n + o(\varepsilon_n)),$$

then

$$\frac{(AS)_n-\mu}{s_n-\mu} = \frac{\rho_n(\rho_{n+1}-\rho_n)}{(\rho_{n+1}\rho_n-2\rho_n+1)} \approx \frac{c_1\rho}{(1-\rho)^2}\left((\rho-1)\varepsilon_n + o(\varepsilon_n)\right).$$

That is, $|(AS)_n-\mu| = \mathcal{O}(|s_n-\mu|^2)$.

The Aitken extrapolation can accelerate fixed-point iteration solving the root x^* for $f(x)$. Steffensen's method is a root-finding algorithm based on such an acceleration technique, the iteration is given by

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

If f is twice differentiable, then clearly the function $g(x) \approx f'(x)$, which recovers the Newton-Raphson method asymptotically if $f(x) \approx 0$.

Theorem 2.3. *Suppose $f'(x^*) \in (-1, 0)$, then the order of convergence is 2 for Steffensen's method.*

Proof. Let us point out the connection between Aitken extrapolation and Steffensen's method. Denote x_0 as the starting point, the intermediate values $z_1 = x_0 + f(x_0)$ and $z_2 = z_1 + f(z_1)$. The Aitken extrapolation finds

$$\begin{aligned} Ax_0 &= x_0 - \frac{(z_1-x_0)^2}{z_2+x_0-2z_1} = x_0 - \frac{|f(x_0)|^2}{z_1+f(z_1)+x_0-2(x_0+f(x_0))} \\ &= x_0 - \frac{|f(x_0)|^2}{f(z_1)-f(x_0)} = x_0 - \frac{f(x_0)}{g(x_0)} = x_1. \end{aligned}$$

The convergence is linear when the iterates x_0, z_1, z_2 are close to the root x^*. Therefore, using the same derivation as Remark 2.2, we find

$$x_1 - x^* = Ax_0 - x^* = \mathcal{O}(|x_0-x^*|^2).$$

□

Remark 2.4. *It should be noticed that Steffensen's method evaluates* f *twice during each iteration, the same as Newton's method (one for* f *and one for* f'*, although the evaluation of* f' *is more expensive in practice). Each evaluation brings an order of* 1 *of convergence on average. From the viewpoint of efficiency, the secant method is preferred, since it only evaluates* f *once each iteration with an order of* $\frac{1+\sqrt{5}}{2}$ *of convergence.*