Example 1.37. Calculate $y = f(x) = x - \sin x$ for $x \to 0$.

Since $x \approx \sin x$ when $x$ is small, the calculation involves a loss significance. The solution is to use the Taylor series

$$x - \sin x = x - \left( x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \right)$$

$$= \frac{x^3}{3!} - \frac{x^5}{5!} + \frac{x^7}{7!} + \cdots$$

1.3.2 Backward stability and numerical stability

Definition 1.35 (Backward stability). An algorithm $\hat{f}(x)$ for computing $y = f(x)$ is backward stable if its backward error is small for all $x$, i.e.

$$\forall x \in \text{dom}(f), \exists \hat{x} \in \text{dom}(f), \text{s.t.} \quad \hat{f}(x) = f(\hat{x}) \Rightarrow E_{\text{rel}}(\hat{x}) \leq c_2\epsilon u,$$  \quad (1.28)

where $c$ is a small constant.

Definition 1.36. An algorithm $\hat{f}(x_1, x_2)$ for computing $y = f(x_1, x_2)$ is backward stable if

$$\forall (x_1, x_2) \in \text{dom}(f), \exists \hat{x}_1, \hat{x}_2 \text{s.t.} \quad \hat{f}(x_1, x_2) = f(\hat{x}_1, \hat{x}_2) \Rightarrow \begin{cases} E_{\text{rel}}(\hat{x}_1) \leq c_1\epsilon u, \\ E_{\text{rel}}(\hat{x}_2) \leq c_2\epsilon u, \end{cases}$$  \quad (1.29)

where $c_1, c_2$ are two small constants.

Corollary 1.37. For $f(x_1, x_2) = x_1 - x_2$, $x_1, x_2 \in \mathbb{R}(F)$, the algorithm $\hat{f}(x_1, x_2) = \text{fl}(f(x_1) - f(x_2))$ is backward stable.

Proof. We have $\hat{f}(x_1, x_2) = (\text{fl}(x_1) - \text{fl}(x_2))(1 + \delta_3)$ from Theorem 1.27. Then Theorem 1.19 implies

$$\hat{f}(x_1, x_2) = (x_1(1 + \delta_1) - x_2(1 + \delta_2))(1 + \delta_3)$$

$$= x_1(1 + \delta_1 + \delta_2 + \delta_3) - x_2(1 + \delta_2 + \delta_3)$$

Take $\hat{x}_1$ and $\hat{x}_2$ to be the two terms in the above line and we have

$$E_{\text{rel}}(\hat{x}_1) = (\delta_1 + \delta_2 + \delta_3)\delta_1),$$

$$E_{\text{rel}}(\hat{x}_2) = (\delta_2 + \delta_3)\delta_3).$$

Then Definition 1.36 completes the proof. \Box

Remark 1.38. Corollary 1.37 does not hold if the condition $x_1, x_2 \in \mathbb{R}(F)$ fails to hold. Example 1.39 illustrates this.

Example 1.39. For $f(x) = 1 + x$, $x \in \mathbb{R}^+$, the algorithm $\hat{f}(x) = \text{fl}(1.0 + \text{fl}(x))$ is not backward stable.

Proof. We prove a stronger statement that implies the negation of (1.28). For each $x \in (0, \epsilon_u)$, Definition 1.15 yields $\hat{f}(x) = 1.0$. Then $\hat{f}(x) = f(\hat{x})$ implies $\hat{x} = 0$, which further implies $E_{\text{rel}}(\hat{x}) = 1$. \Box

Definition 1.38. An algorithm $\hat{f}(x)$ for computing $y = f(x)$ is stable or numerically stable if

$$\forall x \in \text{dom}(f), \exists \hat{x} \text{s.t.} \begin{cases} \frac{|f(x) - f(\hat{x})|}{f(\hat{x})} \leq c_f\epsilon u, \\ E_{\text{rel}}(\hat{x}) \leq c\epsilon u, \end{cases}$$  \quad (1.30)

where $c_f, c$ are two small constants.

Remark 1.40. By Example 1.39, the condition of backward stability is unnecessarily strong in that it rules out some reasonable algorithms, the errors of which are dominated by rounding errors. Definition 1.38 weakens the condition in Definition 1.35 and hence applies to a wider range of algorithms. However, Definition 1.35 is the one usually adopted for stability analysis of an algorithm, c.f. Definition 1.47.

The plot below is a shift of the dashed arrow of the upper triangular in Remark 1.33.

\begin{itemize}
  \item Example 1.42. For $f(x) = 1 + x$, $x \in \mathbb{R}^+$, the algorithm $\hat{f}(x) = \text{fl}(1.0 + \text{fl}(x))$ is stable.
  \end{itemize}

Proof. If $|x| < \epsilon_u$, then $\hat{f}(x) = 1.0$. Choose $\hat{x} = x$, then

$$f(\hat{x}) - x = f(x)$$

and $\frac{|f(x) - f(\hat{x})|}{f(\hat{x})} = \left| \frac{x}{1+x} \right| < 2\epsilon_u$.

Otherwise $|x| \geq \epsilon_u$. Definitions 1.17 and 1.11 yield $x \in \mathbb{R}(F)$. By Theorem 1.19, $\hat{f}(x) = (1 + x(1 + \delta_1))(1 + \delta_2)$, i.e. $\hat{f}(x) = 1 + \delta_2 + x(1 + \delta_1 + \delta_2 + \delta_1 \delta_2)$, where $\delta_1, \delta_2 < \epsilon_u$.

Choose $\hat{x} = x(1 + \delta_1 + \delta_2 + \delta_1 \delta_2)$ and we have

$$E_{\text{rel}}(\hat{x}) = \delta_1 + \delta_2 + \delta_1 \delta_2 < 3\epsilon_u,$$

and we have

$$\frac{|f(x) - f(\hat{x})|}{f(\hat{x})} = \left| \frac{\delta_2}{1 + x(1 + \delta_1 + \delta_2 + \delta_1 \delta_2)} \right| \leq \epsilon_u. $$

where the denominator is never close to zero since $x > 0$. \Box

Remark 1.43. The notions of accuracy and stability qualitatively describes the sensitivity of the answer to small changes of the input for an algorithm. The concept of a condition number quantify these notions.