Lecture 5: Numerical Differentiation

- **Differentiation （微分） and integration （积分）** form the backbone of the mathematical techniques required to describe and analyze physical systems.
- These two mathematical concepts describe how certain quantities(确定的变量) of interest change with respect to either space and time or both (时间域空间域).
- Understanding how to evaluate these quantities numerically is essential to understanding systems beyond the scope of analytic methods.

Numerical Differentiation(数值微分)

Given a set of data or a function, it may be useful to differentiate the quantity (变量) considered in order to determine a physically relevant property (确定其相关物理属性). For instance, given a set of data which represents the position of a particle (质点) as a function of time, then the derivative (一阶导) and second derivative (二阶导) give the velocity (速度) and acceleration (加速度) respectively. From calculus (微积分), the **definition of the derivative** is given by

$$\frac{df(t)}{dt} = \lim_{\Delta t \to 0} \frac{f(t+\Delta t) - f(t)}{\Delta t}$$

Since the derivative is the slope (斜率), the formula on the right is nothing more than a rise-over-run formula for the slope (求取直线上升率的公式). The general idea of calculus is that as $\Delta t \to 0$, then the rise-over-run gives the instantaneous (瞬时的) slope. Numerically, this means that if we take Δt sufficiently small, than the approximation should be fairly accurate. To quantify (量化) and control the error associated with approximating the derivative (近似微分), we make use of **Taylor series expansions (泰勒级数展开)**.

To see how the Taylor expansions are useful, consider the following two Taylor series:

$$f(t+\Delta t)=f(t)+\Delta t\frac{df(t)}{dt}+\frac{\Delta t^2}{2!}\frac{d^2f(t)}{dt^2}+\frac{\Delta t^3}{3!}\frac{d^3f(c_1)}{dt^3}$$
$$f(t-\Delta t)=f(t)-\Delta t\frac{df(t)}{dt}+\frac{\Delta t^2}{2!}\frac{d^2f(t)}{dt^2}-\frac{\Delta t^3}{3!}\frac{d^3f(c_2)}{dt^3}$$

where $c_1\in[t,t+\Delta t]$ and $c_2\in[t,t-\Delta t]$. Subtracting these two expressions gives

$$f(t+\Delta t)-f(t-\Delta t)=2\Delta t\frac{df(t)}{dt}+\frac{\Delta t^3}{3!}\left(\frac{d^3f(c_1)}{dt^3}+\frac{d^3f(c_2)}{dt^3}\right)$$

$$x_0=t \qquad\qquad x_0=t$$
$$x=t+\Delta t \qquad\qquad x=t-\Delta t$$
$$x-x_0=\Delta t \qquad\qquad x-x_0=-\Delta t$$

$$f(x)=\frac{f(x_0)}{0!}+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\ldots$$

By using the **mean-value theorem of calculus (微积分的中值定理)**, we find $f'''(c) = \left(f'''(c_1) + f'''(c_2)\right)/2$. Upon dividing the above expression by $2\Delta t$ and rearranging, we find the following expression for the **first derivative (一阶导数)**

$$\frac{df(t)}{dt} = \frac{f(t+\Delta t) - f(t-\Delta t)}{2\Delta t} - \frac{\Delta t^2}{6}\frac{d^3 f(c)}{dt^3}$$

where the last term is the truncation error (截断误差) associated with the approximation of the first derivative using this particular Taylor series generated expression.

Note that the truncation error in this case is $O\left(\Delta t^2\right)$.

Figure 4.1 demonstrates graphically the numerical procedure and approximation for the second-order slope formula. Here, nearest neighbor points are used to calculate the slope.

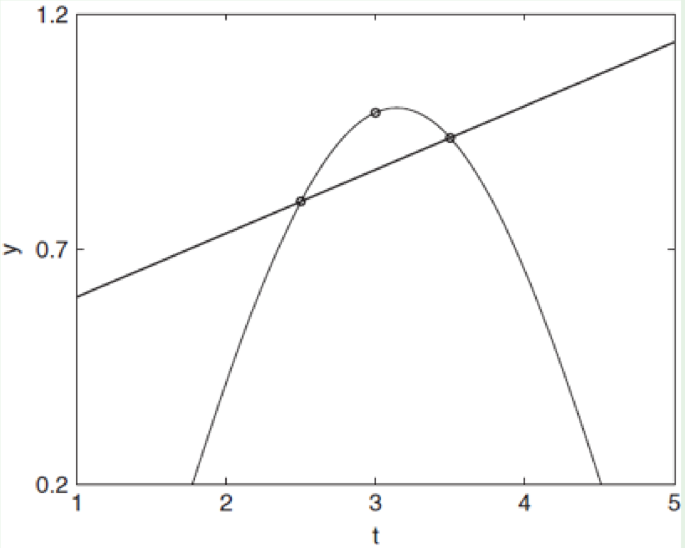

Figure 4.1: Graphical representation of the second-order accurate method for calculating the derivative with finite differences. The slope is simply rise over run where the nearest neighbors are used to determine both quantities. The specific function considered is $y = -\cos(t)$ with the derivative being calculated at $t = 3$ with $\Delta t = 0.5$.

We could improve on this by continuing our Taylor expansion and **truncating (截断) it at higher orders** in Δt. This would lead to higher accuracy schemes. Specifically, by truncating at $O\left(\Delta t^5\right)$, we would have

$$
\begin{aligned}
f(t+\Delta t) = f(t) + \Delta t\frac{df(t)}{dt} + \frac{\Delta t^2}{2!}\frac{d^2 f(t)}{dt^2} \\
+ \frac{\Delta t^3}{3!}\frac{d^3 f(t)}{dt^3} + \frac{\Delta t^4}{4!}\frac{d^4 f(t)}{dt^4} + \frac{\Delta t^5}{5!}\frac{d^5 f\left(c_1\right)}{dt^5} \\
f(t-\Delta t) = f(t) - \Delta t\frac{df(t)}{dt} + \frac{\Delta t^2}{2!}\frac{d^2 f(t)}{dt^2} \\
- \frac{\Delta t^3}{3!}\frac{d^3 f(t)}{dt^3} + \frac{\Delta t^4}{4!}\frac{d^4 f(t)}{dt^4} - \frac{\Delta t^5}{5!}\frac{d^5 f\left(c_2\right)}{dt^5}
\end{aligned}
$$

where $c_1 \in [t, t+\Delta t]$ and $c_2 \in [t, t-\Delta t]$

Again subtracting these two expressions gives

$$f(t+\Delta t)-f(t-\Delta t)=2\Delta t\frac{df(t)}{dt}+\frac{2\Delta t^3}{3!}\frac{d^3f(t)}{dt^3}+\frac{\Delta t^5}{5!}\left(\frac{d^5f(c1)}{dt^5}+\frac{d^5f(c2)}{dt^5}\right)$$

In this approximation, there is third derivative term (三阶导数项) left over which needs to be removed. By using two additional points to approximate the derivative, this term can be removed. Thus we use the two additional points $f(t+2\Delta t)$ and $f(t-2\Delta t)$. Upon replacing Δt by $2\Delta t$ in (4.1.6), we find

$$f(t+2\Delta t)-f(t-2\Delta t)=4\Delta t\frac{df(t)}{dt}+\frac{16\Delta t^3}{3!}\frac{d^3f(t)}{dt^3}+\frac{32\Delta t^5}{5!}\left(\frac{d^5f(c3)}{dt^5}+\frac{d^5f(c4)}{dt^5}\right)$$

where $c_3\in[t,t+2\Delta t]$ and $c_4\in[t,t-2\Delta t]$

By multiplying (4.1.6) by eight and subtracting (4.1.7) and using the mean-value theorem on the truncation terms twice, we find the expression:

$$\frac{df(t)}{dt} = \frac{-f(t+2\Delta t) + 8f(t+\Delta t) - 8f(t-\Delta t) + f(t-2\Delta t)}{12\Delta t} + \frac{\Delta t^4}{30} f^{(5)}(c)$$

where $f^{(5)}$ is the fifth derivative and the truncation is of $O(\Delta t^4)$.

Approximating higher derivatives works in a similar fashion. By starting with the pair of equations (4.1.2) and adding, this gives the result

$$f(t+\Delta t)+f(t-\Delta t)=2f(t)+\Delta t^2\frac{d^2 f(t)}{dt^2}+\frac{\Delta t^4}{4!}\left(\frac{d^4 f(c_1)}{dt^4}+\frac{d^4 f(c_2)}{dt^4}\right)$$

By rearranging and solving for the second derivative, the $O\left(\Delta t^2\right)$ accurate expression is derived

$$\frac{d^2 f(t)}{dt^2}=\frac{f(t+\Delta t)-2f(t)+f(t-\Delta t)}{\Delta t^2}+O\left(\Delta t^2\right)$$

where the truncation error is of $O\left(\Delta t^2\right)$ and is found again by the mean-value theorem to be $-\left(\Delta t^2/12\right)f''''(c)$

This process can be continued to **find any arbitrary derivative**.

Thus, we could also approximate the third, fourth, and higher derivatives using this technique.

It is also possible to **generate backward and forward difference schemes (前向后项微分方案)** by using points only behind or in front of the current point respectively.

Tables 4-6 **summarize the second-order and fourthorder central difference schemes** along with the forward- and backward-difference formulas which are accurate to second-order.

Tbble 4.1 Second-order accurate center-difference formulas

$O\left(\Delta t^2\right)$ center-difference schemes.
$f'(t) = [f(t+\Delta t) - f(t-\Delta t)]/2\Delta t$
$f''(t) = [f(t+\Delta t) - 2f(t) + f(t-\Delta t)]/\Delta t^2$
$f'''(t) = [f(t+2\Delta t) - 2f(t+\Delta t) + 2f(t-\Delta t) - f(t-2\Delta t)]/2\Delta t^3$
$f''''(t) = [f(t+2\Delta t) - 4f(t+\Delta t) + 6f(t) - 4f(t-\Delta t) + f(t-2\Delta t)]/\Delta t^4$

Tbble 4.2 Fourth-order accurate center-difference formulasformulas

$O\left(\Delta t^4\right)$ center-difference schemes.

$$\begin{aligned}
f'(t) =&[-f(t+2\Delta t)+8f(t+\Delta t)-8f(t-\Delta t)+f(t-2\Delta t)]/12\Delta t\\
f''(t) =&[-f(t+2\Delta t)+16f(t+\Delta t)-30f(t)+16f(t-\Delta t)-f(t-2\Delta t)]/12\Delta t^2\\
f'''(t) =&[-f(t+3\Delta t)+8f(t+2\Delta t)-13f(t+\Delta t)+13f(t-\Delta t)-8f(t-2\Delta t)+f(t-3\Delta t)]/8\Delta t^3\\
f''''(t) =&[-f(t+3\Delta t)+12f(t+2\Delta t)-39f(t+\Delta t)+56f(t)-39f(t-\Delta t)+12f(t-2\Delta t)\\
&-f(t-3\Delta t)]/6\Delta t^4
\end{aligned}$$

Tbble 4.3 Second-order accurate forward- and backward-difference formulas

$O\left(\Delta t^2\right)$ forward- and backward-difference schemes.

$$\begin{aligned}
f'(t) &= [-3f(t)+4f(t+\Delta t)-f(t+2\Delta t)]/2\Delta t\\
f'(t) &= [3f(t)-4f(t-\Delta t)+f(t-2\Delta t)]/2\Delta t\\
f''(t) &= [2f(t)-5f(t+\Delta t)+4f(t+2\Delta t)-f(t+3\Delta t)]/\Delta t^2\\
f''(t) &= [2f(t)-5f(t-\Delta t)+4f(t-2\Delta t)-f(t-3\Delta t)]/\Delta t^2
\end{aligned}$$

A final remark is in order concerning these differentiation schemes (关注这些微分方案的阶). The central difference schemes (中心差分方案) are an excellent method for generating the values of the derivative in the interior points of a data set (数据结合内部点). However, **at the end points**, **forward and backward difference methods must be used since they do not have neighboring points** to the left and right respectively. Thus special care must be taken at the end points of any computational domain.

It may be tempting to deduce (推断) from the difference formulas (微分公式) that as $\Delta t \to 0$, the accuracy only improves in these computational methods. However, this line of reasoning (这种推理方式) completely neglects the second source of error in evaluating derivatives: numerical round-off (数值舍入).

例 6.2 设 $f(x)=\cos(x)$。

(a) 利用公式(3)和公式(10)，步长分别为 $h=0.1, 0.01, 0.001$ 和 0.0001，计算 $f'(0.8)$ 的近似值。精度为小数点后 9 位。

(b) 与真实值 $f'(0.8)=-\sin(0.8)$ 进行比较。

解：(a) 设 $h=0.01$，根据公式(3)，可得

$O(\Delta t^2)$

$$f'(0.8)\approx\frac{f(0.81)-f(0.79)}{0.02}\approx\frac{0.689498433-0.703845316}{0.02}\approx-0.717344150$$

设 $h=0.01$，根据公式(10)，可得

$O(\Delta t^4)$

$$\begin{aligned}f'(0.8)&\approx\frac{-f(0.82)+8f(0.81)-8f(0.79)+f(0.78)}{0.12}\\&\approx\frac{-0.682221207+8(0.689498433)-8(0.703845316)+0.710913538}{0.12}\\&\approx-0.717356108\end{aligned}$$

(b) 公式(3)和公式(10)的近似值误差分别为 -0.000011941 和 0.000000017。在本例中，当 $h=0.01$ 时，公式(10)给出的 $f'(0.8)$ 的近似值比公式(3)给出的更好。通过对本例的误差分析可以得出上面的结论。其他的计算如表 6.2 所示。 ■

表 6.2 根据公式(3)和公式(10)得到的数值微分

步长	公式(3)的近似值	公式(3)的误差	公式(10)的近似值	公式(10)的误差
0.1	−0.716161095	−0.001194996	−0.717353703	−0.000002389
0.01	−0.717344150	−0.000011941	−0.717356108	0.000000017
0.001	−0.717356000	−0.000000091	−0.717356167	0.000000076
0.0001	−0.717360000	−0.000003909	−0.717360833	0.000004742

Round-off and optimal step-size(数值舍入和最佳步长)

An unavoidable consequence of working with numerical computations is **round-off error (舍入误差)**. When working with most computations, double precision (双倍精度) numbers are used. This allows for 16-digit accuracy (16位精度) in the representation of a given number. This round-off has significant impact upon numerical computations and the issue of time-stepping(时间步长).

As an example of the impact of round-off, we consider the approximation to the derivative

$$\frac{dy}{dt} \approx \frac{y(t+\Delta t) - y(t)}{\Delta t} + \epsilon(y(t), \Delta t)$$

Where $\epsilon(y(t), \Delta t)$ measures the truncation error (截断误差).

Upon evaluating this expression in the computer, **round-off error occurs** so that

$$y(t+\Delta t) = Y(t+\Delta t) + e(t+\Delta t)$$
$$y(t-\Delta t) = Y(t-\Delta t) + e(t-\Delta t)$$

where *Y (t)* is the approximated value given by the computer and *e(t)* measures the error from the true value *y(t)*.

Thus the combined error between the round-off and truncation gives the following expression for the derivative:

$$\frac{dy}{dt} = \frac{Y(t+\Delta t) - Y(t-\Delta t)}{2\Delta t} + E(y(t), \Delta t)$$

where the total error, *E*, is the combination of round-off and truncation such that

$$E = E_{\text{round}} + E_{\text{trunc}} = \frac{e(t+\Delta t) - e(t-\Delta t)}{2\Delta t} - \frac{\Delta t^2}{6}\frac{d^3 y(c)}{dt^3}$$

We now **determine the maximum size of the error**. In particular, we can bound (限制) the maximum value of round-off and the derivate to be

$$
\begin{aligned}
|e(t+\Delta t)| &\leq e_r \\
|-e(t-\Delta t)| &\leq e_r \\
M &= \max_{c\in[t_n,t_{n+1}]}\left\{\left|\frac{d^3y(c)}{dt^3}\right|\right\}
\end{aligned}
$$

This then gives the **maximum error** to be

$$|E| \leq \frac{e_r+e_r}{2\Delta t}+\frac{\Delta t^2}{6}M=\frac{e_r}{\Delta t}+\frac{\Delta t^2 M}{6}$$

Note that as Δt gets large, **the error grows quadratically (平方地) due to the truncation error**. However, as Δt decreases to zero, **the error is dominated (占主导地位的) by round-off** which grows like $1/\Delta t$.

The error as a function of the step-size Δt for this second-order scheme is represented in Fig. 4.2. Note the dominance (支配地位) of the error on numerical round-off as $\Delta t \to 0$.

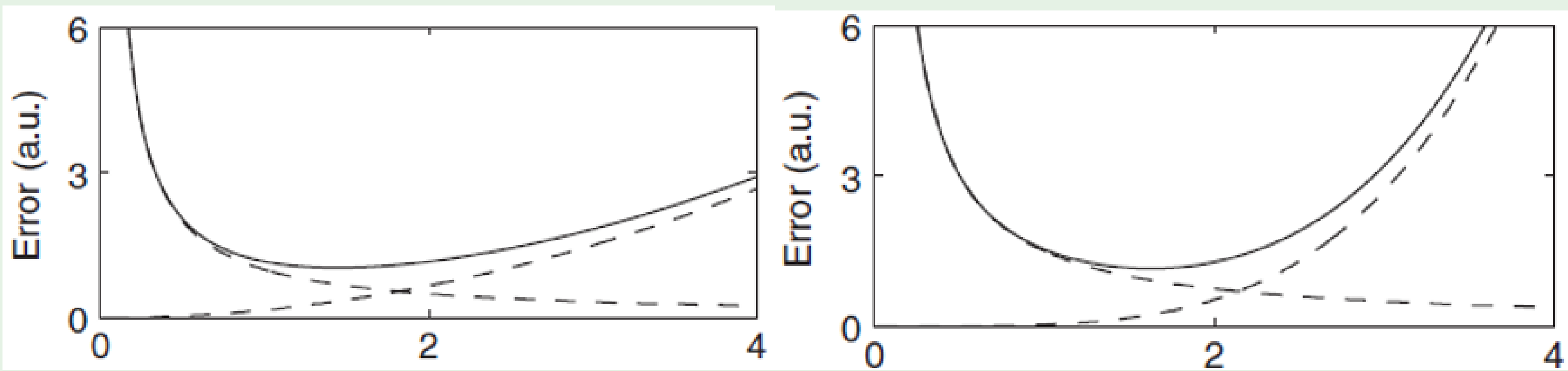

Figure 4.2: Graphical representation of the error which is made up of two components (dotted lines): numerical round-off and truncation. The total error in arbitrary units is shown for both a second-order scheme (top panel) and a fourth-order scheme (bottom panel). For convenience, we have taken $M = 1$ and $e_r = 1$. Note the dominance of numerical round-off as $\Delta t \to 0$.

To **minimize the error**, we require that $\partial|E|/\partial(\Delta t) = 0$. Calculating this **derivative** gives

$$\frac{\partial|E|}{\partial(\Delta t)} = -\frac{e_r}{\Delta t^2} + \frac{\Delta t M}{3} = 0$$

so that

$$\Delta t = \left(\frac{3e_r}{M}\right)^{1/3}$$

This gives the step size resulting in a minimum error.

Thus the smallest step-size is not necessarily the most accurate. Rather, a balance between round-off error and truncation error is achieved to obtain the optimal step-size. For $e_r \approx 10^{-16}$ the optimal $\Delta t \approx 10^{-5}$ Below this value of Δt , numerical round-off begins to dominate the error.

A similar procedure can be carried out for evaluating the optimal step size associated with the $O\left(\Delta t^4\right)$ accurate scheme for the first derivative. In this case

$$\frac{dy}{dt} = \frac{-f(t+2\Delta t) + 8f(t+\Delta t) - 8f(t-\Delta t) + f(t-2\Delta t)}{12\Delta t} + E(y(t), \Delta t)$$

where the total error, *E*, is the combination of round-off and truncation such that

$$E = \frac{-e(t+2\Delta t) + 8e(t+\Delta t) - 8e(t-\Delta t) + e(t-2\Delta t)}{12\Delta t} + \frac{\Delta t^4}{30}\frac{d^5 y(c)}{dt^5}$$

We now determine the maximum size of the error. In particular, we can bound the maximum value of round-off to *e* as before and set $M = \max\{|y'''''(c)|\}$. This then gives the maximum error to be

$$|E| = \frac{3e_r}{2\Delta t} + \frac{\Delta t^4 M}{30}$$

Note that as Δt gets large, the error grows like a quartic (四次函数) due to the truncation error. However, as Δt decreases to zero, the error is again dominated by round-off which grows like $1/\Delta t$.

To minimize the error, we require that $\partial|E|/\partial(\Delta t) = 0$. Calculating this derivative gives

$$\Delta t = \left(\frac{45e_r}{4M}\right)^{1/5}$$

Thus in this case, the optimal step $\Delta t \approx 10^{-3}$ This shows that the error can be quickly dominated by numerical round-off if one is not careful to take this significant effect into account.