Ch. 1. Euler's Method + Related Ideas

Key Points:
- 1st order ODE \( \dot{\varphi} = \dot{\varphi}(t, \varphi) \)
  \((IVP)\) \( \varphi(t_0) = \varphi_0 \)

- Euler's Method
  - basic idea
  - various derivations (differential form, integral form)

- Simple example \[ \begin{cases} \dot{y} = 3y \\ y(0) = 1 \end{cases} \]

- More general conditions - Lipschitz condition
- convergence proof (general case)

- error - local/global

- Definition of "Method of order p"

- Trapezoid Method (Implicit Method)
  - integral derivation
  - error
  - convergence proof

- Theta Method
  - Euler, Trapezoid, Backward Euler
  - error

- Modified Euler - Heun's Method
  - error

- Simpson's Method
Let's start by considering a simple one-dimensional first order initial value problem.

\[
\begin{align*}
\frac{dy}{dt} &= f(t, y) \\
y(t_0) &= y_0
\end{align*}
\]

where \( f \) is sufficiently smooth—more on this later...

We'd like to solve this problem—find \( y(t) \)—up to some final time \( t_f \) ... numerically at a discrete set of points.

**Plan:** use "slope" information to advance a small distance in time

\[
y_1 = y_0 + \left( t_1 - t_0 \right) \cdot f(t_0, y_0) \quad \text{(gives } y_1 \text{)}
\]

then

\[
y_2 = y_1 + \left( t_2 - t_1 \right) \cdot f(t_1, y_1) \quad \text{(gives } y_2 \text{)}
\]

in general

\[
y_{k+1} = y_k + \frac{t_{k+1} - t_k}{h_k} f(t_k, y_k)
\]

**Euler's method**

\[
y_{k+1} = y_k + h_k f(t_k, y_k) \quad k=1, 2, 3, \ldots \quad \text{start with } y_0
\]

explicit formula to "advance" in time.
One way to view this is... rewriting

\[
\frac{y_{k+1} - y_k}{h_k} = f(t_k, y_k)
\]

Essentially we approximate \( \frac{dy}{dt} = \lim_{h \to 0} \frac{y(t+h) - y(t)}{h} \)

approximate the derivative \( \approx \frac{y_{k+1} - y_k}{h_k} \)

where \( h_k \) is "small"

Another way to view this is an "integral, or quadrature, problem"

\[
y(t) = y(t_k) + \int_{t_k}^{t_{k+1}} f(t, y(t)) \, dt \quad \text{(integrate } y' = f(t, y) \text{)}
\]

Approximate the integral (area) ... try left-endpoint approx.

\[
y_{k+1} = y_k + \frac{(t_{k+1} - t_k) \cdot f(t_k, y_k)}{h_k}
\]

Euler's Method

We'll revisit this quadrature approach again later...
Let's examine a simple example.

**EXAMPLE**

Solve \[
\begin{cases}
y' = 3y \\
y(0) = 1
\end{cases}
\]

using Euler's Method, on \(0 \leq t \leq t_F\).

Set up a grid (uniform spacing for simplicity):

\[
t_0, t_1, t_2, \ldots, t_n
\]

\[
t = 0 \quad h = \frac{t_F}{n}
\]

**Euler**

\[y_{k+1} = y_k + h f(t_k, y_k)\]

\[y_{k+1} = y_k + h \cdot 3y_k = (1 + 3h)y_k\]

So

\[y_1 = (1 + 3h)y_0 = (1 + 3h)\]

\[y_2 = (1 + 3h)y_1 = (1 + 3h)^2\]

\[\vdots\]

\[y_n = (1 + 3h)y_{n-1} = (1 + 3h)^n = (1 + \frac{3t_F}{n})^n\]

Note: \(\lim_{n \to \infty} \left(1 + \frac{a}{n}\right)^n = e^a\)

So

\[\lim_{n \to \infty} y_n = \lim_{n \to \infty} \left(1 + \frac{3t_F}{n}\right)^n = e^{3t_F}\]

So \(y_n\) converges to the exact solution \(y(t) = e^{3t}\).

*Normally we can't solve these explicitly, but this was a nice, easy, linear problem.*
Sticking to 2D still...

**Error in Euler's method** ... two sorts of error to think about:

- **Global error** $\varepsilon_k = y_k - y(t_k)$
- **Local error** $\varepsilon_k = y_k - y_{k-1}(t_k)$

$y_{k-1}(t_k)$: solution of ODE passing through previous point

Taylor Expand $y(t_{k+1})$ and compare with Euler:

$y(t_{k+1}) \approx y(t_k) + y'(t_k)h + \frac{1}{2} h^2 y''(t_k) + \ldots$

$y_{k+1} = y_k + h f(t_k, y_k)$

$y(t_{k+1}) - y_{k+1} = y(t_k) - y_k + h [y'(t_k) - f(t_k, y_k)] + \frac{1}{2} h^2 y''(t_k) + \ldots$

$= y(t_k) - y_k + h [f(t_k, y(t_k)) - f(t_k, y_k)] + \frac{1}{2} h^2 y''(t_k) + \ldots$
\begin{align*}
  f(t_k, y(t_k)) & = f(t_k, y_k) + \frac{df}{dy} (t_k, y_k) (y(t_k) - y_k) + \ldots \\
  y(t_{k+1}) - y_{k+1} & = y(t_k) - y_k + h \left[ \frac{df}{dy} (t_k, y_k) (y(t_k) - y_k) \right] + \frac{1}{2} h^2 y''(t_k) + \ldots \\
  y(t_{k+1}) - y_{k+1} & = \left[ 1 + h \frac{df}{dy} (t_k, y_k) \right] [y(t_k) - y_k] + \frac{1}{2} h^2 y''(t_k) + \ldots
\end{align*}

This is the error generated if the solution at the previous step is exact.

We also see the familiar (?) condition for stability of the numerical scheme ... i.e.,

\[ |1 + h \frac{df}{dy}| \leq 1 \]

Expect that global error \( \sim n \cdot h^2 \) \( \sim (\frac{1}{h}) h^2 \sim O(h) \)

Let's see if we can make this a little more rigorous.

+ Generalize for \( \begin{cases} \bar{y}' = \bar{y}(t, \bar{y}) \\ \bar{y}(t_0) = \bar{y}_0 \end{cases} \)
So consider...

\[
\begin{align*}
\vec{y}' &= \vec{f}(t, \vec{y}) & t > t_0 \\
\vec{y}(t_0) &= \vec{y}_0
\end{align*}
\]

\(\vec{y}, \vec{y}_0 \in \mathbb{R}^d\) \(\vec{f} : [t_0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d\) Euclidean Space

For \(\vec{f}\) we require that it satisfy the Lipshitz condition:

\[
\|\vec{f}(t, \vec{x}) - \vec{f}(t, \vec{y})\| \leq \lambda \|\vec{x} - \vec{y}\|
\]

for all \(\vec{x}, \vec{y} \in \mathbb{R}^d\), \(t \geq t_0\), \(\lambda = \text{constant indep. of } \vec{x}, \vec{y}\)

\(\text{i.e. there exists } \lambda > 0 \text{ s.t. this condition holds}\)

Note: We'll often end up requiring that \(\vec{f}\) has a Taylor Series about every point \((t_0, \vec{y})\) \(\Rightarrow\) analytic \(\Rightarrow\) has a positive radius of convergence.

- one might conceptually associate \(\lambda\) with \(\frac{\partial \vec{f}}{\partial \vec{y}}\)
  (or some measure thereof...)

\[\frac{\text{Euler's Method (constant step size for simplicity)}}{\text{}}\]

\[
\vec{y}_{k+1} = \vec{y}_k + h \vec{f}(t_k, \vec{y}_k)
\]
We saw for a simple 1D example that the solution to Euler's method applied to \( y' = 3y \) converged to the exact solution.

More generally, a method is said to be convergent if

\[
\lim_{h \to 0} \left( \max_{k} \left\| y_k - y(t_k) \right\| \right) = 0
\]

for every ODE with Lipschitz function \( \tilde{f} \)

(see text, p. 6 ... slightly streamlined notation)

\( y_k \) is understood to depend on \( h \).

Let's revisit Euler's Method and show

**Thm 1.1** Euler's Method is convergent.

To show this...

Taylor Expand Exact Sol: \( \tilde{y}(t_{k+1}) \approx \tilde{y}(t_k) + h \tilde{y}'(t_k) + \frac{h^2}{2} \tilde{y}''(t_k) + O(h^3) \)

\[
\tilde{y}_{k+1} = \tilde{y}_k + h \tilde{f}(t_k, \tilde{y}_k)
\]

\[
\tilde{y}(t_{k+1}) - \tilde{y}_{k+1} = \tilde{y}(t_k) - \tilde{y}_k + h \left[ \tilde{f}(t_k, \tilde{y}(t_k)) - \tilde{f}(t_k, \tilde{y}_k) \right] + O(h^2)
\]

\[
= \tilde{e}_{k+1} - \tilde{e}_k
\]

(note: Iserles uses different sign) but no worries, norms are coming

\[
\tilde{e}_{k+1} = \tilde{e}_k + h \left[ \tilde{f}(t_k, \tilde{y}(t_k)) - \tilde{f}(t_K, \tilde{y}(t_k) - \tilde{e}_k) \right] + O(h^2)
\]
Taking norms...

\[ \| \tilde{e}_{k+1} \| = \| \tilde{e}_k + h \left[ \tilde{f}(t_k, \tilde{x}(t_k)) - \tilde{f}(t_k, \tilde{x}(t_k) - \tilde{e}_k) \right] + o(h) \]

triangle inequality

Lipschitz cond.

\[ \leq \| \tilde{e}_k \| + h \| \tilde{f}(t_k, \tilde{x}(t_k)) - \tilde{f}(t_k, \tilde{x}(t_k) - \tilde{e}_k) \| + o(h) \]

\[ \leq \| \tilde{e}_k \| + h \lambda \| \tilde{x}(t_k) - (\tilde{x}(t_k) - \tilde{e}_k) \| + o(h) \]

\[ = \| \tilde{e}_k \| + h \lambda \| \tilde{e}_k \| + \frac{o(h)}{c h^2} \]

so

\[ \| \tilde{e}_{k+1} \| \leq (1 + h \lambda) \| \tilde{e}_k \| + \frac{o(h)}{c h^2} \]

For convergence, we'd like to show \( \| e_k \| \to 0 \) as \( h \to 0 \).

\[ \| e_k \| \leq c h^2 \quad \text{note: } \| e_0 \| = 0 \]

\[ \| e_2 \| \leq \frac{(1 + h \lambda)^2 (c h^2) + (c h^2)}{c h^2 \left[ 1 + (1 + h \lambda) \right]} \]

\[ \| e_3 \| \leq \frac{(1 + h \lambda) c h^2 \left[ 1 + (1 + h \lambda) \right] + c h^2}{c h^2 \left[ 1 + (1 + h \lambda) + (1 + h \lambda)^2 \right]} \]

\[ \vdots \]

\[ \| e_k \| \leq c h^2 \left[ 1 + (1 + h \lambda) + (1 + h \lambda)^2 + \ldots + (1 + h \lambda)^{k-1} \right] \]
but \( S = 1 + p + p^2 + \ldots + p^{k-1} \)

\[
\frac{pS}{S-pS} = 1 - p^k
\]

\[
S = \frac{1 - p^k}{1 - p}
\]

\[\|e_k\| \leq c h^2 \left[ \frac{1 - (1+h\lambda)^k}{1 - (1+h\lambda)} \right] = ch^2 \left[ \frac{1 - (1+h\lambda)^k}{-h\lambda} \right]
\]

\[\|e_k\| \leq c h \left[ \frac{(1+h\lambda)^k - 1}{\lambda} \right]
\]

This proves that Euler's Method is convergent (i.e. \( \|e_k\| \to 0 \) as \( h \to 0 \))

and that the global error decays as \( O(h) \).

Since \( 1+h\lambda < e^{h\lambda} \) (e.g. \( e^{h\lambda} = 1 + h\lambda + \frac{1}{2} (h\lambda)^2 + \ldots \))

\( (1+h\lambda)^n < e^{h\lambda n} \) but \( hn = T_F \) (i.e. \( h = \frac{T_F}{n} \))

\[\|e_k\| \leq ch \left[ e^{\lambda T_F} - 1 \right]
\]

\[\|e_k\| \leq \frac{ch}{\lambda} \left[ e^{\lambda T_F} - 1 \right]
\]
General Method of Order \( p \)

Consider

\[ Y_{k+1} = \overline{Y}_k \left( \hat{x}, h, \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k \right) \]

single-step method

or multi-step method

We say this method is order \( p \) [i.e. local error \( \sim O(h^{p+1}) \)] if when we equate \( Y_{k+1} \) with \( y(t_{k+1}) \) and \( Y_k \) with \( y(t_k) \), etc.

we find

\[ y(t_{k+1}) - \overline{Y}_k \left( \hat{x}, h, \hat{y}(t_0), \hat{y}(t_1), \ldots, \hat{y}(t_k) \right) = O(h^{p+1}) \]

recall our calculations with \( \epsilon_k \) and \( \delta_k \)

if we set \( y(t_0) = y_k \) the "new" error is just the local error.
Other Methods

Recall our integral version of the derivation of Euler's Method.

\[ y(t_{k+1}) = y(t_k) + \int_{t_k}^{t_{k+1}} f(t, y) \, dt \]

\[ f(t, y(t_k)) \]

\[ t_k \quad t_{k+1} \]

A better approximation of the area under the curve between \( t_k \) and \( t_{k+1} \) ... is the Trapezoid method

\[ y(t_{k+1}) \approx y(t_k) + \frac{1}{2} (t_{k+1} - t_k) \left[ f(t_k, y(t_k)) + f(t_{k+1}, y(t_{k+1})) \right] \]

So the Trapezoid Method is ... (for \( \tilde{y}' = \tilde{f}(t, \tilde{y}) \))

\[ \tilde{y}_{k+1} = \tilde{y}_k + \frac{1}{2} h \left[ \tilde{f}(t_k, \tilde{y}_k) + \tilde{f}(t_{k+1}, \tilde{y}_{k+1}) \right] \]

Observations:

- This is not an explicit method (implicit)

Q: What order is this method?
Q: Is the trapezoid rule convergent?

MORE ON THIS LATER!
Order of Trapezoid method (characterizing the local error)

start with

\[ \tilde{y}_{k+1} = \tilde{y}_k + \frac{1}{2} h \left[ f(t_k, \tilde{y}_k) + f(t_k, \tilde{y}_{k+1}) \right] \]

+ plug in exact solution (use \( \tilde{y}' = f(t, \tilde{y}) \))

\[ Q \equiv \tilde{y}(t_{k+1}) = \left\{ \tilde{y}(t_k) + \frac{1}{2} h \left[ \tilde{y}'(t_k) + \tilde{y}'(t_{k+1}) \right] \right\} \]

\[ Q = \tilde{y}(t_k) + \tilde{y}'(t_k) h + \frac{1}{2} h^2 y''(t_k) + \frac{1}{6} h^3 y'''(t_k) + O(h^4) \]

\[ = \left\{ \tilde{y}(t_k) + \frac{1}{2} h \left[ \tilde{y}'(t_k) + (\tilde{y}'(t_k) + h y''(t_k) + \frac{1}{2} h^2 y'''(t_k)) \right] \right\} \]

\[ Q = O(h^3) \text{ i.e. local error looks like } O(h^3) \]

so \( \boxed{\text{Trapezoid Method is of order 2}} \)

We suspect that the global error is \( O(h^2) \) and that the Trapezoid Method converges but let's prove this result...
Thm 1.2 The Trapezoid rule is convergent

To show this... let \( \tilde{e}_{k+1} = \tilde{y}(t_{k+1}) - \tilde{y}_{k+1}, \quad \tilde{e}_k = \tilde{y}(t_k) - \tilde{y}_k \)

**Taylor**
\[
\tilde{y}(t_{k+1}) = \tilde{y}(t_k) + h \tilde{y}'(t_k) + \frac{1}{2} h^2 \tilde{y}''(t_k) + \frac{1}{3!} h^3 \tilde{y}'''(t_k) + o(h^3)
\]

**Trapezoid**
\[
\tilde{y}_{k+1} = \tilde{y}_k + \frac{1}{2} h \left[ f(t_k, \tilde{y}_k) + f(t_{k+1}, \tilde{y}_{k+1}) \right]
\]

Subtract:
\[
y(t_{k+1}) - \tilde{y}_{k+1} = \tilde{y}(t_k) - \tilde{y}_k
\]

We could use this approach if we assume \( f \) is differentiable so that we can deal with the \( \tilde{y}'' \) term.

However, let's use the result we just found, namely
\[
y(t_{k+1}) = y(t_k) + \frac{1}{2} h \left[ f(t_k, y(t_k)) + f(t_{k+1}, y(t_{k+1})) \right] + o(h^3).
\]
\[ y(t_{k+1}) = y(t_k) + \frac{h}{2} \left[ f(t_k, y(t_k)) + f(t_{k+1}, y(t_{k+1})) \right] - O(h^2) \]

\[ \tilde{y}_{k+1} = \tilde{y}_k + \frac{h}{2} \left[ f(t_k, \tilde{y}_k) - f(t_{k+1}, \tilde{y}_{k+1}) \right] \]

\[ \| \tilde{e}_{k+1} \| = \| \tilde{e}_k + \frac{h}{2} \left[ f(t_k, y(t_k)) - f(t_k, \tilde{y}_k) \right] \| + O(h^3) \]

\[ \leq \| \tilde{e}_k \| + \frac{h}{2} \| f(t_k, y(t_k)) - f(t_k, \tilde{y}_k) \| \]

\[ \leq \| \tilde{e}_k \| + \frac{h}{2} \| f(t_k, \tilde{y}_k) - f(t_{k+1}, \tilde{y}_{k+1}) \| + C h^3 \]

\[ \leq \| \tilde{e}_k \| + \frac{h}{2} \| \tilde{e}_k \| + \frac{h}{2} \| \tilde{e}_{k+1} \| + C h^3 \]

\[ (1 - \frac{1}{2} h \lambda) \| \tilde{e}_{k+1} \| \leq (1 + \frac{1}{2} h \lambda) \| \tilde{e}_k \| + C h^3 \]

(assuming \( 1 - \frac{1}{2} h \lambda > 0 \))

\[ \| \tilde{e}_{k+1} \| \leq \frac{1 + \frac{1}{2} h \lambda}{1 - \frac{1}{2} h \lambda} \| \tilde{e}_k \| + \frac{C h^3}{(1 - \frac{1}{2} h \lambda)} \]

(compare with Euler: \( \| \tilde{e}_{k+1} \| \leq (1 - h \lambda) \| \tilde{e}_{k} \| + h^2 \))
Again this has the form
\[ e_{n+1} \leq a \cdot e_n + b \]
\[ e_1 \leq b \]
\[ e_2 \leq a \cdot b + b = (a+1) \cdot b \]
\[ e_3 \leq a \cdot (a+1) \cdot b + b = (a^2+a+1) \cdot b \]
\[ e_k \leq (a^{n-1} + \ldots + a+1) \cdot b \]
\[ \frac{1-a^k}{1-a} \]

So
\[ \|e_k\| \leq \frac{c \cdot h^3}{(1-\frac{3}{2}h \lambda)} \left[ 1 - \left( \frac{1+\frac{1}{2}h \lambda x^k}{1-\frac{1}{2}h \lambda} \right) \right] \leq \frac{c \cdot h^3 \left[ 1 - \left( \frac{1+\frac{1}{2}h \lambda x}{1-\frac{1}{2}h \lambda} \right)^k \right]}{(1-\frac{1}{2}h \lambda) - (x+\frac{1}{2}h \lambda)} \]

\[ \|e_k\| \leq \frac{c \cdot h^2}{\lambda} \left[ \left( \frac{1+\frac{1}{2}h \lambda}{1-\frac{1}{2}h \lambda} \right)^k - 1 \right] \]

one can argue ... see book (Iserles, p. 9+10) that this term is \( O(h^2) \) as \( h \to 0 \).

Therefore, trapezoid method converges and its associated global error is \( O(h^2) \)* one order better than Euler!
Other Methods

\[ y(t_{k+1}) = y(t_k) + \int_{t_k}^{t_{k+1}} f(t,y) \, dt \]

- Approximate based on right endpoint

\[ y_{k+1} = y_k + hf(t_{k+1}, y_{k+1}) \]
- This is **Backward Euler - Implicit**
- First order accurate
- Converges

Midpoint rule

\[ y_{k+1} = y_k + hf(t_k + \frac{1}{2}h, \frac{1}{2}(y_k + y_{k+1})) \]
- **Midpoint Rule - Implicit**
- Second order accurate
- Converges

See HW Exercise 1.1
Henri’s Method (Modified Euler)

This method is based on the trapezoidal method but is modified to make it an explicit method.

\[
\text{Trapezoidal: } \quad y_{k+1} = y_k + \frac{1}{2} h \left[ f(t_k, y_k) + f(t_{k+1}, y_{k+1}) \right]
\]

\[
\text{Henri’s: } \quad y_{k+1} = y_k + \frac{1}{2} h \left[ K_1 + K_2 \right]
\]

where

\[
K_1 = f(t_k, y_k)
\]

\[
K_2 = f(t_{k+1}, y_k + hf(t_k, y_k))
\]

\(\text{effectively an Euler approximation of } y_{k+1}\).

Comments on Henri’s Method

- This is an explicit method
- Verify order of accuracy
- Verify convergence

\(2^{\text{nd}}\) order
Theta Method

\[ y_{k+1} = y_k + h \left[ \theta f(t_k, y_k) + (1-\theta) f(t_{k+1}, y_{k+1}) \right] \]

Comments:
\[ \theta = 1 : \text{ Euler (explicit)} \]
\[ \theta = \frac{1}{2} : \text{ Trapezoid (implicit)} \]
\[ \theta = 0 : \text{ Backward Euler (implicit)} \]
\[ \theta = \text{ other} \ldots \text{ provides flexibility ... extra tuning parameter} \]

The text shows that (see p.14)

\[ \mathcal{E} \]
\[ y(t_{k+1}) - y(t_k) - h \left[ \theta f(t_k, y(t_k)) + (1-\theta) f(t_{k+1}, y(t_{k+1})) \right] \]

\[ = (\theta - \frac{1}{2}) h^2 y''(t_k) + \left( \frac{1}{2} \theta - \frac{1}{6} \right) h^3 y'''(t_k) + O(h^4) \]

In principle, the best term to knock out of the local error is the \( O(h^2) \) term. However, in practice there may be situations where knocking out the \( O(h^2) \) term makes the actual error lower \( (\theta = \frac{2}{3}) \).