

lecture 23

Consider IVP

$$\underline{x}'(t) = F(\underline{x}) \quad t \in [0, T]$$

$$\underline{x}(0) = \underline{x}_0$$

with $F \in C^1(\mathbb{R}^d, \mathbb{R}^d)$.

$$\left(\begin{array}{l} \text{This covers} \\ \underline{y}' = F(\underline{y}, t) \quad \text{by writing} \\ \underline{x}(t) = \begin{bmatrix} t \\ \underline{y} \end{bmatrix} \text{ and } \underline{x}' = \begin{bmatrix} 1 \\ F(\underline{x}) \end{bmatrix} =: \overline{F}(\underline{x}) \end{array} \right)$$

Notation numerical solution

$$\text{mesh} \quad h = \frac{T}{N} \quad , \quad t_n = nh$$

$$\text{Num. solution } \overline{\underline{x}}_n \approx \underline{x}(t_n)$$

One step method:

$$\overline{\underline{x}}_{n+1} = G(\overline{\underline{x}}_n, \overline{\underline{x}}_{n+1}, h) \quad \text{with } \overline{\underline{x}}_0 = \underline{x}_0.$$

If Φ depends on $\bar{X}_{n+1}$, then we call it an implicit method, otherwise, the method is explicit.

Foundation for numerical methods:

$$\bar{X}(t_{n+1}) = \bar{X}(t_n) + \underbrace{\int_{t_n}^{t_{n+1}} F(\bar{X}(s)) ds}_I$$

Quadrature approximations

• $I \approx F(\bar{X}(t_n)) h$ leads to

forward Euler

$$\bar{X}_{n+1} = \bar{X}_n + F(\bar{X}_n) h = G(\bar{X}_n, h)$$

• $I \approx F(\bar{X}(t_{n+1})) h$ leads to implicit Euler

$$\bar{X}_{n+1} = \bar{X}_n + F(\bar{X}_{n+1}) h = G(\bar{X}_n, \bar{X}_{n+1}, h)$$

$$\bullet I \approx \frac{F(\bar{X}(t_n)) + F(\bar{X}(t_{n+1}))}{2} \cdot h$$

leads to Crank-Nicolson method

$$\bar{X}_{n+1} = \bar{X}_n + \frac{F(\bar{X}_n) + F(\bar{X}_{n+1})}{2} h$$

etc.

Ex 1 $X' = aX,$
 $X(0) = X_0$

Forward Euler: $\bar{X}_{n+1} = \bar{X}_n + f(\bar{X}_n) h$
 $= (1 + ah) \bar{X}_n$

$$\Rightarrow \bar{X}_n = (1 + ah)^n X_0 \approx e^{at_n} X_0$$

Implicit Euler:

$$\bar{X}_{n+1} = \bar{X}_n + f(\bar{X}_{n+1}) h = \bar{X}_n + a\bar{X}_{n+1} h$$

$$\Rightarrow \bar{X}_{n+1} = \frac{\bar{X}_n}{1 - ah} \quad (\text{assuming } ah \neq 1)$$

$$\text{and } \bar{X}_n = \frac{x_0}{(1-ah)^n} \approx e^{atn} x_0$$

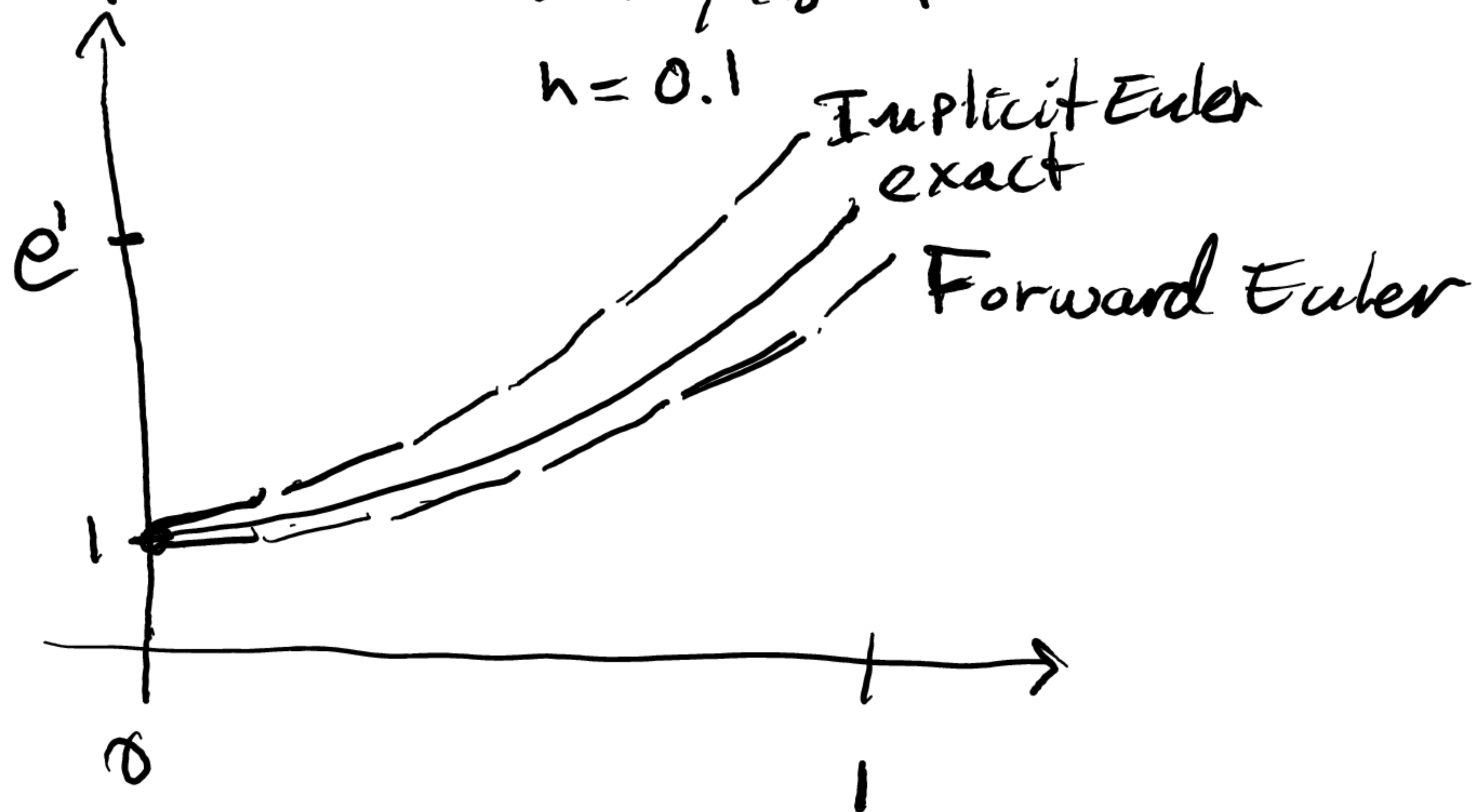
CN method

$$\bar{X}_{n+1} = \bar{X}_n + a \frac{\bar{X}_n + \bar{X}_{n+1}}{2} h$$

$$\Rightarrow \bar{X}_{n+1} = \frac{\left(1 + \frac{ah}{2}\right) \bar{X}_n}{1 - \frac{ah}{2}} \quad \left(\text{assuming } ah \neq 2\right)$$

$$\Rightarrow \bar{X}_n = \frac{\left(1 + \frac{ah}{2}\right)^n}{\left(1 - \frac{ah}{2}\right)^n} x_0 \approx e^{atn} x_0$$

Rough sketch
of methods $a=1$, $x_0=1$



Ex 2

For $\bar{x}' = A\bar{x}$

$$\bar{x}(0) = \bar{x}_0,$$

Forward Euler becomes

$$\bar{x}_{n+1} = \bar{x}_n + F(\bar{x}_n)h$$

$$= (I + hA)\bar{x}_n$$

$$\Rightarrow \bar{x}_n = (I + hA)^n \bar{x}_0$$

Implicit Euler

$$\bar{x}_{n+1} = \bar{x}_n + F(\bar{x}_{n+1})h$$

$$= \bar{x}_n + (A\bar{x}_{n+1})h$$

$$\Rightarrow (I - Ah)\bar{x}_{n+1} = \bar{x}_n$$

$$\bar{x}_{n+1} = (I - Ah)^{-1} \bar{x}_n \quad \left(\begin{array}{l} h \text{ sufficiently} \\ \text{small so that} \\ I - Ah \text{ invertible} \end{array} \right)$$

and $\bar{X}_n = (I - Ah)^{-n} X_0$.

An explicit method $\bar{X}_{n+1} = G(\bar{X}_n, h)$ of course has a unique solution for every $\bar{X}_n$ and h ; implicit methods typically too if $h > 0$ is small:

Lemma 1

If $\exists L > 0$ s.t. $|F(X) - F(Y)| \leq L|X - Y|$ for all $X, Y \in \mathbb{R}^d$, then the implicit Euler method

$$\bar{X}_{n+1} = \bar{X}_n + F(\bar{X}_{n+1})h$$

has a unique solution $\bar{X}_{n+1}$ for each $\bar{X}_n \in \mathbb{R}^d$ and $h < \frac{1}{L}$.

Proof Let $\psi(Y) := \bar{X}_n + F(Y)h$.

Then $|\psi(X) - \psi(Y)| \leq |F(X) - F(Y)|h \leq hL|X - Y|$.

Let $I_0 := \psi(\bar{X}_n)$ & $I_{j+1} = \psi(I_j)$ $j = 0, 1, \dots$

Then

$$|\mathcal{I}_{j+1} - \mathcal{I}_j| \leq hL |\mathcal{I}_j - \mathcal{I}_{j-1}| \leq \underbrace{(hL)}_{<1}^j |\mathcal{I}_1 - \mathcal{I}_0|$$

$\Rightarrow \{\mathcal{I}_j\} \subset \mathbb{R}^d$ is Cauchy in $(\mathbb{R}^d, |\cdot|)$

and $\mathcal{I} := \lim_{j \rightarrow \infty} \mathcal{I}_j = \lim_{j \rightarrow \infty} \mathcal{I}_{j+1} = \lim_{j \rightarrow \infty} \psi(\mathcal{I}_j)$

$$= \psi(\mathcal{I}).$$

This means $\bar{\mathcal{X}}_{n+1} = \mathcal{I}$ solves

$$\bar{\mathcal{X}}_{n+1} = \mathcal{I} = \psi(\mathcal{I}) = \bar{\mathcal{X}}_n + F(\mathcal{I})h = \bar{\mathcal{X}}_n + F(\bar{\mathcal{X}}_{n+1})h.$$

Uniqueness: If also $Z = \psi(Z)$, then

$$|\mathcal{I} - Z| \leq |\psi(\mathcal{I}) - \psi(Z)| \leq \underbrace{hL}_{<1} |\mathcal{I} - Z|;$$

a contradiction



Remark Lemma 1 extends to CN for

$$h < \frac{2}{L}$$

Error estimates

For $\bar{X}_{n+1} = G(\bar{X}_n, \bar{X}_{n+1}, h)$ solving (1),
we seek to derive a bound

$$|\bar{X}_n - X(T)| \leq Ch^p \text{ for small } h = \frac{T}{N}$$

where $C > 0$ depends on X_0, T, F and the numerical method. $p > 0$ is called the order of the method.

Assumptions For any sufficiently small $h = \frac{T}{N}$ (e.g. large $N \in \mathbb{N}$), it holds that

A1 $\exists!$ solution Z to $Z = G(Z, Z, h)$
(see Lemma 1)

A2 (Stability) There exist $c_1, c_2 > 0$
s.t. $|\bar{X}_n| \leq c_2 |X_0| + c_1$ for all
 $n \in \{0, 1, \dots, N\}$

A.3 (Consistency) The local error

$$\tau(\bar{x}_n) := \Phi_h(\bar{x}_n) - \bar{\Phi}_h(\bar{x}_n) (= \bar{X}(h; \bar{x}_n) - \bar{x}_{n+1})$$

satisfies $\max_{n \in \{0, 1, \dots, N\}} |\tau(\bar{x}_n)| \leq C_3 h^{p+1}$

for some $p > 0$ and $C_3 > 0$.

Comments

- The alternative error $\tau(\bar{X}(t_n))$, along the exact solution, is called the local truncation error
- If $\lim_{h \downarrow 0} \frac{\tau(\bar{X})}{h} = 0$ for any $\bar{X} \in \mathbb{R}^d$, then $\bar{X}$ is called consistent.

Lemma 2 If $F \in C^1(\mathbb{R}^d, \mathbb{R}^d)$

is globally Lipschitz, then

Forward Euler and Implicit Euler both satisfy A1, A2 and A3 with $p = 1$.

Proof forward Euler

Let $L > 0$ be the Lipschitz constant of F .

[A1] Holds as $\bar{X}_{n+1} = \bar{X}_n + h F(\bar{X}_n)$ is explicit. (see Lemma 1 for implicit Euler)

[A2] As $|F(\bar{X}_n)| \leq |F(\bar{X}_n) - F(0)| + |F(0)|$
 $\leq L |\bar{X}_n - 0| + |F(0)|,$

we get that

$$\begin{aligned} |\bar{X}_{n+1}| &\leq |\bar{X}_n| + |F(\bar{X}_n)| h \\ &\leq (1 + Lh) |\bar{X}_n| + |F(0)| h \end{aligned}$$

$$\leq (1 + Lh)^2 |\bar{X}_{n-1}| + \left(\sum_{i=0}^1 (1 + Lh)^i \right) |F(0)| h$$

$$\leq (1 + Lh)^{n+1} |\bar{X}_0| + \left(\sum_{i=0}^n (1 + Lh)^i \right) |F(0)| h$$

using that $1 + Lh \leq e^{Lh}$, $n+1 \leq N$
 $Nh = T$, we get

$$|\tilde{X}_n| \leq \underbrace{e^{LT}}_{C_1} |\tilde{X}_0| + \underbrace{T e^{LT}}_{\leq \hat{C}} |F(0)|.$$

A3 Forward Euler discrete flow:

$$\Phi_h(\bar{X}_n) = \bar{X}_n + h F(\bar{X}_n)$$

Exact sol:

$$\phi_h(\bar{X}_n) = \bar{X}_n + \int_0^h F(X(s; \bar{X}_n)) ds$$

$$\tau(\bar{X}_n) = \Phi_h(\bar{X}_n) - \phi_h(\bar{X}_n)$$

$$= \int_0^h F(X(s; \bar{X}_n)) ds - F(\bar{X}_n) h$$

$$= \int_0^h F(X(s; \bar{X}_n)) - F(\bar{X}_n) ds$$

$$= \int_0^h \int_0^s \frac{d}{dt} F(X(t; \bar{X}_n)) dt ds$$

$$= \int_0^h \int_0^s DF(X(t; \bar{X}_n)) F(X(t; \bar{X}_n)) dt ds$$

$$|\tau(\bar{X}_n)| \leq \int_0^h \int_0^s \underbrace{\|DF(X(t; \bar{X}_n))\|_2}_{\substack{F \text{ Lipschitz} \leq L}} \underbrace{\|F(X(t; \bar{X}_n))\|}_{\leq \hat{C}} dt ds$$

$$\leq L \hat{C} \int_0^h \int_0^s dt ds = \underbrace{L \hat{C}}_{\leq 2} h^2$$

where F globally Lipschitz $\Rightarrow$

$$|X(t; \bar{X}_n)| \stackrel{(*)}{\leq} e^{Lt} (|\bar{X}_n| + |F(0)|)$$

$$\stackrel{A.2}{\leq} e^{LT} (C_2 |X_0| + e^{LT} C_1 + |F(0)|) = R$$

$$\Rightarrow |F(X(t; \bar{X}_n))| \leq \max_{\overline{B(0, R)}} |F(X)| =: \hat{C}$$

and $\sup_{\mathbb{R}^d} \|\nabla F(X)\|_2 \leq L.$

⊙ Exercise 1 week for above inequality

Theorem 1

Let $F \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ be globally Lipschitz and let Assumptions A1, A2, and A3 hold for a numerical method for some $p > 0$.

$$\text{Then } \max_{n \in \{0, \dots, N\}} |\bar{X}_n - X(t_n)| \leq Ch^p$$

for a $C > 0$ that depends on F, T, X_0 and Φ .

Note • By Lemma 2, we have $p=1$

for forward Euler & Implicit Euler.

• One can show that $p=2$ for CN.

Proof We only consider $n=N$.

(similar argument holds for any $n < N$)

$$|\bar{X}_N - X(t_N)| = |\Phi(0, \bar{X}_N) - \Phi(t_N, X_0)|$$

$$= \left| \sum_{j=1}^N \Phi(t_{N-j}, \bar{X}_j) - \Phi(t_{N-j-1}, \bar{X}_{j-1}) \right|$$

$$\leq \sum_{j=1}^N |\phi(t_{N-j}, \bar{X}_j) - \phi(t_{N+1-j}, \bar{X}_{j-1})|$$

$$\boxed{\phi_{t_{N+1-j}} = \phi_{t_{N-j}} \circ \phi_h}$$

$$= \sum_{j=1}^N |\phi(t_{N-j}, \bar{X}_j) - \phi(t_{N-j}, \phi_h(\bar{X}_{j-1}))|$$

$$\leq \sum_{j=1}^N e^{LT} \underbrace{|\bar{X}_j - \phi_h(\bar{X}_{j-1})|}_{\tau(\bar{X}_{j-1})}$$

$$A^3 \leq N e^{LT} \cdot C_3 h^{p+1} = \underbrace{(hN)}_{=T} \cdot C_3 h^p$$

$$\text{as } |\phi(t, \bar{X}) - \phi(t, \bar{Y})| \leq e^{Lt} |\bar{X} - \bar{Y}|$$

by Theorem 1 in Lecture 17,

where L is F 's Lipschitz constant. ▀