Chapter 2

Reviews on Partial Differential Equations and Difference Equations

1. Properties of PDEs

In this chapter, we study the key defining properties of partial differential equations (PDEs). First of all, there are more than one 'independent' variables $t, x, y, z, \ldots$. Associated to these is so called a 'dependent' variable u (of course there could be more than one dependent variables) which is a function of those independent variables,

$$u = u(t, x, y, z, \ldots) \tag{2.1}$$

We now provide a bunch of basic definitions and examples on PDEs.

Definition: A PDE is a relation between the independent variables and the dependent variable u via the partial derivatives of u.

Definition: The order of PDE is the highest derivative that appears.

Example: $F(x, y, u, u_x, u_y) = 0$ is the most general form of first-order PDE in two independent variables x and y.

Example: $F(t, x, y, u, u_t, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0$ is the most general form of second-order PDE in three independent variables t, x and y.

Example: $u_t - u_{xx} = 0$ is a second-order PDE in two independent variables t and x.

Example: $u_{xxxx} + (u_y)^3 = 0$ is a fourth-order PDE in two independent variables x and y.

Definition: $\mathcal{L}$ is called a linear operator if $\mathcal{L}(u+v) = \mathcal{L}u+\mathcal{L}v$ for any functions u and v.

Definition: A PDE $\mathcal{L}u = 0$ is called a linear PDE if $\mathcal{L}$ is a linear derivative operator.

Definition: A PDE $\mathcal{L}u = g$ is called an inhomogeneous linear PDE if $\mathcal{L}$ is a linear derivative operator and if $g \neq 0$ is a given function of the independent variables. If $g = 0$, it is called a homogeneous linear PDE.

Example: The following PDEs are homogeneous linear:
$u_t + u_x = 0$ (transport); $u_t + xu_x = 0$ (transport); $u_{xx} + u_{yy} = 0$ (Laplace's equation)

Example: The following PDEs are homogeneous nonlinear:
$u_t + uu_x = 0$ (Burgers' equation with shock wave); $u_{tt} - u_{xx} + u^3 = 0$ (wave with interaction); $u_t + uu_x + u_{xxx} = 0$ (dispersive wave);

Example: The following PDEs are inhomogeneous linear:
$\cos(xy^2)u_x - y^2u_y = \tan(x^2 + y^2)$

Remark: In general, we reserve t for the temporal variable, and x, y and z for the three spatial variables in modeling PDEs for fluid dynamics, physical phenomena, etc. in the usual sense, i.e., three spatial dimension with one time dimension.

2. Well-posedness of PDEs

When solving PDEs, one often encounters a problem that has more than one solution (non-uniqueness) if few auxiliary conditions are imposed. Then the problem is called underdetermined. On the other hand, if too many conditions are given, there may be no solution at all (non-existence) and in this case, the problem is overdetermined.

The well-posedness property of PDEs is therefore required in order for us to enable to solve the given PDE system successfully. Well-posed PDEs of proper initial and boundary conditions follows the following fundamental properties:

1. Existence: There exists at least one solution $u(x, t)$ satisfying all these conditions,

2. Uniqueness: There is at most one solution,

3. Stability: The unique solution $u(x, t)$ depends in a stable manner on the data of the problem. This means that if the data are changed a little, the corresponding solution changes only a little as well.

3. Classifications of Second-order PDEs

PDEs arise in a number of physical phenomena to describe their natures. Some of the most popular types of such problems include fluid flows, heat transfer, solid mechanics and biological processes. These types of equations often fall into one of three types, (i) hyperbolic PDEs that are associated with advection, (ii) parabolic PDEs that are most commonly associated with diffusion, and (iii) elliptic PDEs that most commonly describe steady states of either parabolic or hyperbolic problems.

In reality, not many problems fall simply into *one* of these three types, rather most of them involve combined types, e.g., advection-diffusion problems. Mathematically, however, we can rather easily determine the type of a general second-order PDEs, which we are going to briefly discuss here.

In general, let's consider the PDE of form with nonzero constants a_{11}, a_{12}, and a_{22}:

$$a_{11}u_{xx} + 2a_{12}u_{xy} + a_{22}u_{yy} + a_1u_x + a_2u_y + a_0u = 0, \tag{2.2}$$

which is a second-order linear equation in two independent variables x and y with six constant coefficients.

Theorem: By a linear transformation of the independent variables, the equation can be reduced to one of three forms:

1. Elliptic PDE: if $a_{12}^2 < a_{11}a_{22}$, it is reducible to

$$u_{xx} + u_{yy} + L.O.T = 0 \tag{2.3}$$

   where $L.O.T$ denotes all the lower order terms (first or zeroth order terms).

2. Hyperbolic PDE: if $a_{12}^2 > a_{11}a_{22}$, it is reducible to

$$u_{xx} - u_{yy} + L.O.T = 0 \tag{2.4}$$

3. Parabolic PDE: if $a_{12}^2 = a_{11}a_{22}$ (the condition for parabolic is in between those of elliptic and hyperbolic), it is reducible to

$$u_{xx} + L.O.T = 0 \tag{2.5}$$

Remark: Notice the similarity between the above classification and the one in analytic geometry. We know from analytic geometry that, given (again assuming nonzero constants a_{11}, a_{12}, and a_{22})

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + a_1x + a_2y + a_0 = 0, \tag{2.6}$$

Then Eq. 2.6 becomes

1. Ellipsoid if $a_{12}^2 < a_{11}a_{22}$

2. Hyperbola if $a_{12}^2 > a_{11}a_{22}$
3. Parabola if $a_{12}^2 = a_{11}a_{22}$.

Note again that parabola is in between ellipsoid and hyperbola. See Fig. 1 for an illustration.

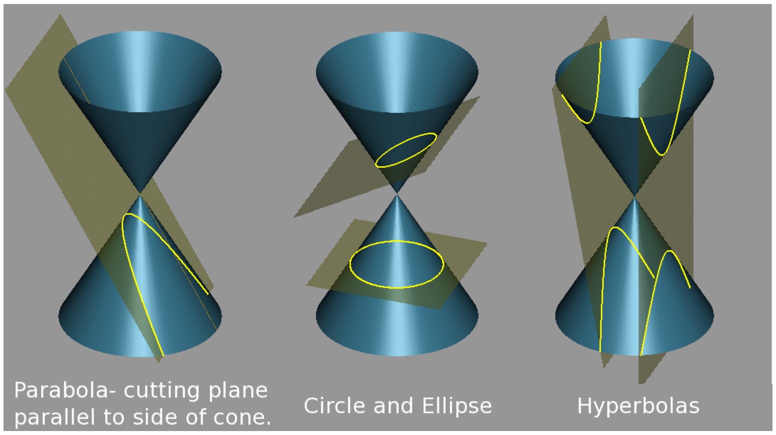

Figure 1. Three major types of conic section from analytic geometry – Image source: Wikipedia

Example: $u_{xx} - 5u_{xy} = 0$ is hyperbolic; $4u_{xx} - 12u_{xy} + 9u_{yy} + u_y = 0$ is parabolic; $4u_{xx} + 6u_{xy} + 9u_{yy} = 0$ is elliptic.

Example: The wave equation is one of the most famous examples in hyperbolic PDEs. We write the wave equation as

$$u_{tt} = c^2 u_{xx} \text{ for } -\infty < x < \infty, c \neq 0. \tag{2.7}$$

Factoring the derivative operator, we get

$$\Big(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\Big)\Big(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\Big)u = 0 \tag{2.8}$$

Considering the characteristic coordinates $\xi = x + ct$ and $\eta = x - ct$, we obtain

$$0 = \Big(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\Big)\Big(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\Big)u = \Big(-2c\frac{\partial}{\partial \xi}\Big)\Big(2c\frac{\partial}{\partial \eta}\Big)u \tag{2.9}$$

Hence, we conclude that the general solution must have a form $u(x,t) = f(x+ct) + g(x-ct)$, the sum of two functions, one (g) is a wave of any shape traveling to the the *right* at speed c, and the other (f) with another arbitrary shape traveling to the the *left* at speed c. We call the two families of lines, $x \pm ct = \mathit{constant}$, the characteristic lines of the wave equation.

Example: One very simple and famous example in the parabolic PDEs is so called the diffusion equation

$$u_t = k u_{xx}, \text{ with } k \text{ constant and } (x,t) \in D \times T \tag{2.10}$$

One of the important properties in the diffusion equations is to have the maximum principle. Recall that the maximum principle says if $u(x,t)$ is the solution of Eq. 2.10 on $D \times T = [x_{min}, x_{max}] \times [T_0, T_1]$ in space-time, then the maximum value of $u(x,t)$ is assumed only on the initial and domain boundary of $D \times T$. That is, the maximum value only occurs either initially at $t = T_0$ or on the sides $x = x_{min}$ or $x = x_{max}$.

Remark: The fundamental properties of the two types of PDEs can be briefly compared in the following table. The physical meanings in Table 1 are also illustrated in Fig. 2 and Fig. 3.

Table 1. Comparison of Waves and Diffusions: Fundamental properties of the wave and diffusion equations are summarized.

Property	Waves	Diffusions
(1) speed of propagation	finite ($\leq c$)	∞
(2) singularities for $t > 0$?	transported along characteristics (with speed $= c$)	lost immediately
(3) well-posed for $t > 0$?	yes	yes (at least for bounded solutions)
(4) well-posed for $t < 0$?	yes	no
(5) maximum principle?	no	yes
(6) behavior as $t \to \infty$	energy is constant so does not decay (i.e., simple advection without diffusion)	decays to zero
(7) information	transported	lost gradually

4. Discretization

We consider the cell-centered (rather than cell interface-centered) notation for discrete cells x_i and the conventional temporal discretization t^n:

$$x_i = (i - \frac{1}{2})\Delta x, i = 1, ..., N, \tag{2.11}$$

$$t^n = n\Delta t, n = 0, ...M. \tag{2.12}$$

Then the cell interface-centered grid points are written using the 'half-integer' indices:

$$x_{i+\frac{1}{2}} = x_i + \frac{\Delta x}{2}. \tag{2.13}$$

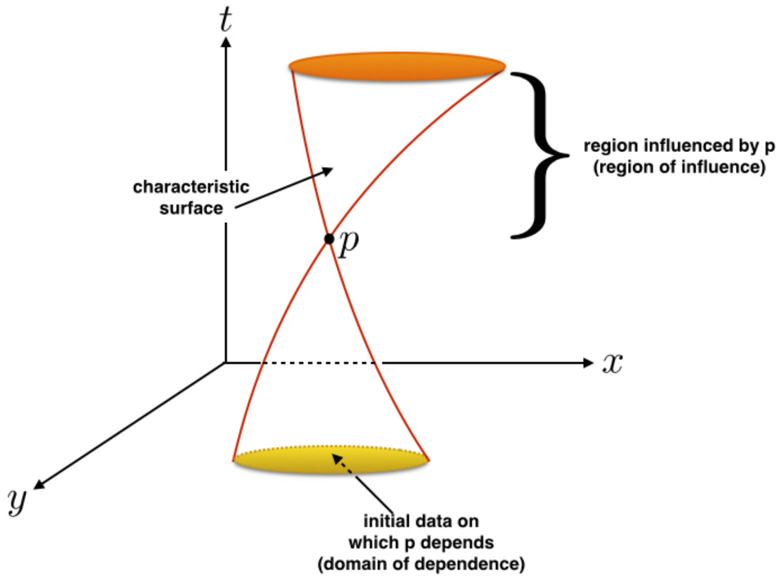
t
x
y
p
characteristic surface
region influenced by p (region of influence)
initial data on which p depends (domain of dependence)

Figure 2. Domain and boundaries for the solution of hyperbolic PDEs in 2D. Note that any information or disturbance introduced at p is going to affect *only* the region called the 'region of influence' but nowhere. Such information is propagated with the finite advection speed along the characteristic surface which forms the conic region of influence. On the other hand, if the characteristic surface can be extended backward in time to the place where the initial data is imposed, this forms another conic section on the lower part of the figure which is called the 'domain of dependence'.

Definition: Let $u_i^n = u(x_i, t^n)$ be the pointwise values of the exact solution of a given PDE at discrete points (x_i, t^n). This is the analytical solution of the PDE and satisfies it without any form of numerical errors.

Definition: Let U_i^n be the numerical approximations to the exact solution of the PDE. For instance, U_i^n represents

$$U_i^n \approx u_i^n \text{ for FDM.} \tag{2.14}$$

Definition: Let D_i^n be the exact solution of the associated 'difference equation (DE)' of the PDE, e.g., the forward in time backward in space (FTBS):

$$\frac{D_i^{n+1} - D_i^n}{\Delta t} = -a\frac{D_i^n - D_{i-1}^n}{\Delta x}. \tag{2.15}$$

Since D_i^n is the exact solution of the DE, there is *no* round-off errors involved. When we study numerical solution of PDEs, the solutions are affected by numerical errors. They mainly come from two sources of numerical errors, and we are now ready to define them.

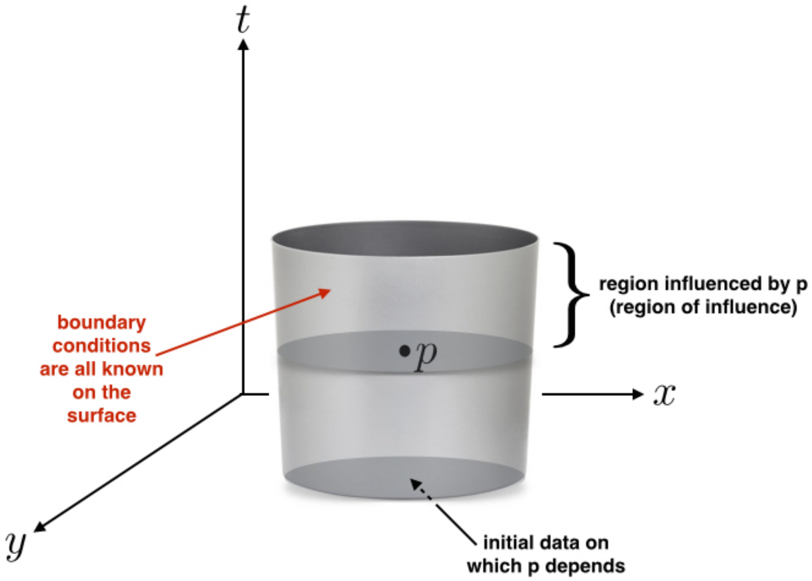
t
x
y
p
boundary conditions are all known on the surface
region influenced by p (region of influence)
initial data on which p depends

Figure 3. Domain and boundaries for the solution of parabolic PDEs in 2D. Note that from a given point p in the mid plane, there is only one physically meaningful direction that is positive in t. Therefore, any information at p influences the entire region onward from p, called the 'region of influence'. Such information can only marches forward in time under the assumption that all boundary conditions around the surface and the initial condition are known.

Definition: The *discretization error* E_d^n at (x_i, t^n) is defined by

$$E_{d,i}^n = u_i^n - D_i^n. \tag{2.16}$$

Definition: The *round-off error* $E_{r,i}^n$ at (x_i, t^n) is defined by

$$E_{r,i}^n = D_i^n - U_i^n. \tag{2.17}$$

Definition: The *global error* $E_{g,i}^n$ at (x_i, t^n) is defined by

$$E_{g,i}^n = u_i^n - U_i^n. \tag{2.18}$$

Note by definition, $E_{g,i}^n = E_{d,i}^n + E_{r,i}^n$.

Definition: We say that the numerical method is *convergent* at t^n in a given norm $||\cdot||$ if

$$\lim_{\Delta x, \Delta t \to 0} ||E_g^n|| = 0. \tag{2.19}$$

Remark: We note that the discretization error $E_{d,i}^n$ is the sum of the truncation error $E^n_{T,i}$ for the DE Eq. 2.15 and any numerical errors $E^n_{B,i}$ introduced by the numerical handling of boundary conditions.

Remark: We define the round-off error $E^n_{r,i}$ by the numerical errors introduced after a repetitive number of arithmetic computer operations in which the computer constantly rounds off the numbers to some significant digits.

5. The Fundamental Theorem of Numerical Methods – The Lax Equivalence Theorem for Linear PDEs

The ultimate goal in this chapter is to show (at least partially) one of the theorems that is very powerful to provide us great levels of insights in numerical differential equations. Briefly speaking, the theorem says, for linear PDEs,

$$\texttt{consistency + (absolute) stability} \Longleftrightarrow \texttt{convergence}$$

Let us take a moment to think about the meaning of this theorem. It says that if the numerical scheme *converges* to a (weak) solution provided the scheme is proven to be consistent (we are going to define it shortly) and stable. So, what is good about it? The good news is that in numerically solving many PDE systems, it is often very difficult to directly show convergence of a given numerical method because not many PDEs have their exact analytical solutions available (see the definition of convergence in Eq. 2.19). Without guaranteeing the existence of such analytical solutions, one cannot possibly say her/his numerical scheme converges to a mathematically meaningful and correct solution at all.

A nice workaround is instead to look at numerical stability and consistency that are based on a recurrence property of the numerical method acting on the discrete grid data. The Lax Equivalence theorem then indicates that such numerical method is indeed a convergent method that produces a well-defined weak solution. Now let's take a look at this nice theorem in more details.

First, we define few more things.

Definition: Let $\mathcal{N}$ be the (linear) numerical operator mapping the approximate solution at one time step to the approximate solution at the next time step. Then a general explicit numerical method can be written as

$$U^{n+1}_i = \mathcal{N}(U^n_i). \tag{2.20}$$

We define the *one-step error* $E^n_{1step,i}$ by

$$E^n_{1step,i} = u^n_i - \mathcal{N}(u^{n-1}_i), \tag{2.21}$$

and the *local truncation error* $E^n_{LT,i}$ by

$$E^n_{LT,i} = \frac{1}{\Delta t} E^n_{1step,i}. \tag{2.22}$$

We have already discussed the *the order of method* previously, and we now can define it again using the local truncation error.

Definition: We say that the numerical method is *of order p (or pth order accurate)* if for all sufficiently smooth data with compact support, the local truncation error is given as

$$E^n_{LT,i} = \mathcal{O}(\Delta t^p + \Delta x^p). \tag{2.23}$$

Remark: One can obviously introduce a method that has different orders of accuracy in space and time, i.e., a method that is of p-th order accurate in time and r-th order accurate in space can be defined as

$$E^n_{LT,i} = \mathcal{O}(\Delta t^p + \Delta x^r). \tag{2.24}$$

In this case, the numerical solution in a fully resolved state – both temporally and spatially – will exhibit its convergence rate dominated by the lower rate between the two, i.e., assuming $\Delta t \approx \Delta x \approx h \ll 1$ the error will be dominated by $h^{\min\{p,r\}}$, or simply

$$E^n_{LT,i} = \max\Big[\mathcal{O}(\Delta t^p), \mathcal{O}(\Delta x^r)\Big]. \tag{2.25}$$

5.1. Consistency

Let's now formally define consistency of the numerical methods.

Definition: We say the numerical method is *consistent* in $||\cdot||$ with the given PDE if

$$\lim_{\Delta t, \Delta x \to 0} ||E^n_{LT}|| = 0 \tag{2.26}$$

for all smooth functions $u(x,t)$ that satisfies the given PDE.

Remark: In words, the numerical consistency is a measure to see if the numerical operator $\mathcal{N}$ is in fact 'consistent' with the PDE of interest in a sense that the method should introduce a small error in any one step.

Remark: On the other hand, the numerical stability is a property that the numerical method does not produce any local errors that grow catastrophically and hence a bound on the global error can be obtained in terms of these local errors.

5.2. Stability Theory

The form of stability bounds in this section provides a useful information in analyzing 'linear' methods. It has to be emphasized that for 'nonlinear' methods, the same technique we adopt for the linear method becomes hard to apply, and therefore one has to provide a different approach to discuss nonlinear stability (these topics will be covered in AMS 260). We limit our interest in the linear stability theory in this chapter.

In order to assess stability of the linear PDEs, we essentially need to bound the global error $E^n_{g,i} = u^n_i - U^n_i$ using a recurrence relation. Applying the *linear* numerical operator $\mathcal{N}$ to U^n_i, we obtain

$$U^{n+1}_i = \mathcal{N}(U^n_i) = \mathcal{N}(u^n_i - E^n_{g,i}). \tag{2.27}$$

The global error at t^{n+1} is now

$$\begin{aligned} E^{n+1}_{g,i} &= u^{n+1}_i - U^{n+1}_i && (2.28)\\ &= u^{n+1}_i - \mathcal{N}(u^n_i - E^n_{g,i}) && (2.29)\\ &= u^{n+1}_i - \mathcal{N}(u^n_i) + \mathcal{N}(E^n_{g,i}) && (2.30)\\ &= \Delta t E^{n+1}_{LT,i} + \mathcal{N}(E^n_{g,i}). && (2.31) \end{aligned}$$

Note that the first term in Eq. 2.31 is the new one-step error introduced in this time step, and this term is therefore related to the consistency control of the numerical method. On the other hand, the second term in the parenthesis is the effect of the numerical method on the *previous* global error $E^n_{g,i}$ and this is the term that is to do with the stability control.

Definition: We say the linear numerical method defined by the linear operator $\mathcal{N}$ is *stable* in $||\cdot||$ if there is a constant C such that

$$||\mathcal{N}^n|| \le C, \quad \forall n\Delta t \le T, \tag{2.32}$$

for each time T.

Note: We note here that the superscript n on $\mathcal{N}$ represents the *nth power* of the matrix (or linear operator) obtained by repeated applications of the linear operator $\mathcal{N}$. This is, however, not true for nonlinear operators.

Remark: In particular, the numerical method is stable if $||\mathcal{N}|| \le 1$, since in this case, we have

$$||\mathcal{N}^n|| \le ||\mathcal{N}||^n \le 1. \tag{2.33}$$

Theorem: The Lax Equivalence Theorem for linear difference methods states that, for a well-posed consistent, linear method, stability is necessary and sufficient for convergence.

A full proof can be found in a book by Richtmyer and Morton, *Difference Methods for Initial-Value Problems*, Wiley-Interscience, 1967, and we only partially prove the sufficient part of the claim:

`consistency + (absolute) stability` $\Longrightarrow$ `convergence`

Proof: We are going to show

$$\lim_{\Delta t, \Delta x \to 0} ||E_g^{n+1}|| = 0. \tag{2.34}$$

Since $\mathcal{N}$ is linear, Eq. 2.31 becomes, recursively,

$$\begin{aligned}
||E_g^{n+1}|| &\leq \Delta t||E_{LT}^{n+1}|| + ||\mathcal{N}(u^n - E_g^n) - \mathcal{N}(u^n)|| && (2.35)\\
&= \Delta t||E_{LT}^{n+1}|| + ||\mathcal{N}(E_g^n)|| && (2.36)\\
&\leq \Delta t||E_{LT}^{n+1}|| + ||\mathcal{N}||||E_g^n|| && (2.37)\\
&\leq \Delta t||E_{LT}^{n+1}|| + C||E_g^n|| && (2.38)\\
&\leq \Delta t||E_{LT}^{n+1}|| + C\Big(||\mathcal{N}||||E_g^{n-1}|| + \Delta t||E_{LT}^n||\Big) && (2.39)\\
&\cdots && (2.40)\\
&\leq \Delta t\sum_{j=1}^{n+1} C^{n+1-j}||E_{LT}^j|| + C^{n+1}||E_g^0|| && (2.41)\\
&\leq \tilde{D}(n+1)\Delta t||E_{LT}|| + \tilde{C}||E_g^0|| && (2.42)\\
&= \tilde{D}t^{n+1}||E_{LT}|| + \tilde{C}||E_g^0||, && (2.43)
\end{aligned}$$

where $||E_{LT}|| = \max_{1\leq j\leq n+1} ||E_{LT}^j||$, and for some $\tilde{C}$ and $\tilde{D}$.

Now if we let $\Delta t, \Delta x \to 0$, then $||E_g^0|| \to 0$, since it is the global error on resolving the discrete initial data. It has to go to zero when the grid gets more and more refined unless the initial data has some numerical error to start with (i.e., ill-posed problems).

Also, if we let $\Delta t \to 0$, then $||E_{LT}|| \to 0$, since the method is consistent by assumption. Therefore, we prove $||E_g^{n+1}|| \to 0$ as $\Delta x, \Delta t \to 0$, and the method is convergent.

□

Note: It is not hard to show that the the sufficient condition also holds when $\mathcal{N}$ is contractive, i.e.,

$$||\mathcal{N}(P) - \mathcal{N}(Q)|| \leq ||P - Q||. \tag{2.44}$$

Remark: One can also say the method is stable in $||\cdot||$ if

$$||U^{n+1}|| \leq ||U^n||, \tag{2.45}$$

for all n. To show this, let us assume Eq. 2.45. Recalling $U^{n+1} = \mathcal{N}(U^n)$, we have

$$\frac{||\mathcal{N}(U^n)||}{||U^n||} = \frac{||U^{n+1}||}{||U^n||} \leq 1, \tag{2.46}$$

for $||U^n|| \neq 0$. Since Eq. 2.46 is true for all n, we can take sup to get

$$\sup_{U \neq 0} \frac{||\mathcal{N}(U)||}{||U||} \leq 1 \tag{2.47}$$

which gives

$$||\mathcal{N}|| \leq 1. \tag{2.48}$$

Hence Eq. 2.45 implies the method is stable.