Functional analysis and aspects of non-linear control theory

by: Neil Carmichael

Submission for degree of Ph.D. at the University of Warwick

Control Theory Centre
University of Warwick

June 1982

Acknowledgements

I would like to record my thanks for Dr. A.J. Pritchard, my supervisor, Dr. M.D. Quinn and Dr. G. Salut for technical discussions. Gratitude is also due to Professor J-P Babary, and L.A.A.S. Toulouse, for their hospitality from October to December (1981). Not least, however, are my debts to a number of friends who have provided support over the past three years.

Summary

This thesis studies the state reconstruction problem for a class of non-linear systems. This class is that of perturbed linear systems. The properties of the linear part are used to arrive at results for the complete system. Whilst this is a common technique in mathematics and physics its use in non-linear infinite dimensional systems theory has not been extensively investigated. The present work makes such an investigation with a view to indicating the successes, and limitations, of such a treatment. As to contribution, as far as the author is aware, many of the results are new both in precise statement and general approach.

Chapter 1 introduces, and motivates, the formulation adopted. Chapter 2 provides some useful information on linear infinite dimensional control theory. Chapter 3 gives, subject to certain, perhaps restrictive, conditions, a rigorous statement, and proof, of the basic theorems. Here, as elsewhere, the standard fixed point results are used. Parts of this chapter are extracts from, as yet unpublished, joint work with A.J. Pritchard and M.D. Quinn. Chapter 4 relaxes some of the conditions in 3 and applies the same techniques to other areas. Chapter 5 surveys, in a formal fashion, the more constructive, numerical aspects of the preceding results with a view to indicating directions for this important area of further research.

It is concluded that the "perturbed linear" approach used here can give results that are both theoretically and computationally useful. The strength of the requirements placed on the linear part, however, indicates a challenging area for future investigations: a constructive approach to intrinsically non-linear problems.

CONTENTS

CHAPTER I : Introduction

1.1 Generalities

Control and observation problems for linear dynamical systems have been widely studied. In reality, many practical problems concern models which are non-linear in nature. It is thus desirable to develop an analysis of control and observation properties which is applicable to non-linear systems; or at least to classes of non-linear systems, the term "non-linear system" being a slight misnomer since it only specifies exclusion from a particular class. Such a theory should lead to constructive information, since the fact that a solution exists does not often solve a significant practical problem. This thesis studies aspects of such a non linear systems theory for a class of systems which might be called "semi-linear". That is to say, they appear as the sum of a linear and a non-linear part.

For such systems the general aim is to use properties of the linear part in establishing formulations which allow one to prove results for the complete system. A simple example of this procedure is afforded by

Example 1.1

Consider $\dot{z} = Az + f(z)$; $z(0) = z_0$ where A is a linear operator and $f(\cdot)$ a non-linearity. Then, under appropriate conditions (see Chapter 3) we can make sense of the variation of constants formulation

$$z(t) = e^{At}z_0 + \int_0^t e^{A(t-s)}f(z(s))ds \qquad \ldots \qquad (1.1)$$

If we take the right hand side of (1.1) to define an operator action on z , $\Phi(z)$, then a fixed point of Φ is regarded as defining a solution of the original, non-linear, problem.

From a strictly mathematical viewpoint this is an antique technique: a complicated problem is approached by regarding it as a perturbation of a simpler one whose analysis is well-developed. Though widely exploited in the mathematical theory of non-linear differential equations such ideas have only been applied in a disjointed and fragmentary fashion to problems arising in control theory. This state of affairs is to be contrasted with the use of manifold and differential geometric techniques in the "geometric" non-linear control theory of Brockett et al.. Some major contrasts are:

a) global v. local : the differential-geometric approach attempts to deduce global information about controllability and observability - the perturbation approach is necessarily local often both in time and in initial data (it may however sometimes be extended to provide global information).

b) "existential" v. constructive : the geometric ideas (as in the theory of dynamical systems proper) provide qualitative information about the solution - they rarely (if ever) lead directly to a numerical solution. By contrast (perhaps due to the influence of Brouwer) fixed point formulations, such as may arise from perturbation techniques, have been the subject of much "constructive" attention.

c) finite v. infinite (dimensional) : whilst there is a theory of infinite dimensional Banach (and Hilbert) manifolds it has yet to be extensively applied to dynamical systems resulting from partial differential equations. The perturbation techniques (perhaps because of their less sophisticated requirements) have been applied to both ordinary and partial differential equations. Indeed some of the results for semi-linear systems have been developed in the first instance specifically for partial differential equations, without regard to their implications for finite dimensional systems.

Thus, in conclusion, simplicity, particularly when it is effective in solving significant problems, should not be despised. In the present work motivation derives from a desire to provide constructive answers to problems arising from semi-linear partial differential equations. Moreover, suitably equipped with a philosophically based pessimism concerning the limitations of applied mathematics we may be prepared to accept only local answers. It should not be surprising that perturbation techniques constitute a suitable approach.

1.2 Formulation

Consider an observed dynamical system given, at least formally, by

$$\dot{z} = f(z,u,t) \quad ; \quad z(0) = z_0$$
$$y = h(z,u,t) \qquad \ldots \qquad (1.2)$$

where u is the input to the system and is assumed known, z is the state, and y denotes the output. Take an initial guess for the state trajectory $\bar{z}(\cdot)$ and let

$$z = \bar{z} + z'$$
$$y = h(\bar{z},u,t) + y' \qquad \ldots \qquad (1.3)$$

substitution in (1.2) gives

$$\dot{z}' = A(t)z' + \bar{f}(z',u,t) + f(\bar{z},u,t) - \dot{\bar{z}}$$
$$z'(0) = z_0' \qquad \ldots \qquad (1.4)$$
$$y' = C(t)z' + \bar{h}(z',u,t)$$

where $z_0' = z_0 - \bar{z}(0)$.

The (time-varying) linear operators $A(t)$, $C(t)$ represent the linear part of the expansion after local approximation about $\bar{z}$; $\bar{f}(\cdot,\cdot,.)$ and $\bar{h}(\cdot,\cdot,\cdot)$ represent the, higher order, non-linear terms of the approximation.

If $\bar{z}$ satisfies the original dynamics with the specified initial condition then we have

$$\dot{\bar{z}} = f(\bar{z},u,t) \; ; \; \bar{z}(0) = \bar{z}_0 \qquad \ldots \qquad (1.5)$$

When (1.5) does not hold $f(\bar{z},u,t) - \dot{\bar{z}}$ is an additional known quantity in the z' equation. Such known quantities, appearing in either the dynamics or the output equation, do not alter the results of this thesis. For ease of exposition such terms are therefore ignored.

Our aim, then, is to use knowledge of the linear theory of state reconstruction to provide an approach to reconstructing the state of the non-linear systems (1.1). Although reconstructing an initial state for the system which is the linear part of (1.4), i.e.

$$\dot{z} = A(t)z' \qquad y = C(t)z' \qquad \ldots \qquad (1.6)$$

has received much attention (see, for example, Chapter 9 of Curtain-Pritchard, [1]) we shall, again for ease of exposition, restrict attention to time-invariant systems written as

$$\dot{z}' = Az' + f(z') \qquad y' = Cz' \qquad \ldots \qquad (1.7)$$

Note that the non-linearity in the output equation has been dropped. This too can be done without any real loss of generality in the methods to be described.

In Chapter 4 we briefly study the (non-linear) problem of joint state and parameter estimation by our methods. In this case the original system has the form

$$\dot{z} = f(z,u,\alpha,t) \; ; \quad z(0) = z_0 \qquad y = h(z,u,\alpha,t) \qquad \ldots \qquad (1.8)$$

where α lies in the chosen parameter space. As before we make a local approximation about a guess $\overline{z}(\cdot)$ (for the state trajectory) and $\overline{\alpha}$ (the parameter). Performing simplifications as above we arrive at

$$\dot{z}' = A\, z' + A_1\alpha' + f(z,\alpha) \qquad y' = C\, z' \qquad \ldots \qquad (1.9)$$

Of course systems such as (1.7) and (1.9) may arise naturally, rather than by the approximation procedure described above. The next section contains some examples.

1.3 Some models

The examples in this thesis are based on the simplest standard models of non-linear (second-order) parabolic and hyperbolic equaions, i.e. the non-linear heat and wave equations respectively. Here we indicate some circumstances in which models of the type (1.7) might arise. These should be taken as providing evidence that the considerations of this thesis are potentially applicable to the real world. (See Henry [1], Reed [1] for this, and other, evidence.)

Example 1.2 (reaction - diffusion)

Models of chemical reactions often give rise to equations of the form

$$\frac{\partial z}{\partial t} + a\frac{\partial z}{\partial x} - \frac{\partial^2 z}{\partial x^2} = f(z) \quad \ldots \qquad (1.10)$$

More precisely, consider $S_1,\ldots,S_N$ being N chemical species which participate in R independent reactions. These reactions take place over a region $\Omega \subset \mathbb{R}^3$. Let c_i be the concentration of S_i then

$$\frac{\partial c_i}{\partial t} = \mathrm{div}\,(D_i \,\mathrm{grad}\, c_i) + \sum_{j=1}^{R} \alpha_{ij} f_j \quad \ldots \qquad (1.11)$$

where the conservation of mass is expressed as $\sum_{i=1}^{N} \alpha_{ij} S_j = 0 \quad (j=1,\ldots,R)$.

D_i represents the diffusion coefficient for the species S_i and $f_j(c_1,\ldots,c_N)$ is the rate of the $j^{\underline{th}}$ reaction.

Example 1.3 (population genetics)

Consideration of the probabilities of genetic events often gives rise to study of "Fisher's equation"

$$\frac{\partial z}{\partial t} = \frac{\partial^2 z}{\partial x^2} + f(z) \qquad x \in \mathbb{R}^1$$
$$z(0) = z_0 \qquad \ldots \qquad (1.12)$$

Such equations are also used to describe the geographic distribution of plant, animals or epidemics.

Example 1.4 (Navier - Stokes)

The physical laws governing the flow of a viscous incompressible fluid yield the model

$$\frac{\partial q}{\partial t} + (q.\nabla)q = \nu\Delta q - \frac{1}{\rho}\,\text{grad } p$$
$$\text{div } q = 0 \qquad \ldots \qquad (1.13)$$

where p is the pressure, q the velocity and ρ,ν are positive constants (representing density and kinematic viscosity respectively).

By a transformation of Kato-Fujita we obtain a model in the desired semi-linear form. More simply one often considers the simplified form known as Burger's equation viz. (see Burgers in Additional References, p.166)

$$\frac{\partial z}{\partial t} + z\frac{\partial z}{\partial x} - \frac{\partial^2 z}{\partial x^2} = 0 \qquad \ldots \qquad (1.14)$$

rather than the completeNavier-Stokes system (1.13). Such equations have been subject to much research - partly due to their possible relation to turbulent behaviour in fluids.

The above examples give rise to semi-linear parabolic equations; in fact, they are semi-linear diffusion equations. Many other processes give rise to such models e.g. non-linear heat conduction, re-distribution of impurities in semi-conductors; see Henry [1] for further details. One characteristic feature of such equations (in contrast to linear diffusion equations) is the possible presence of travelling wave solutions. Though much analysis has been devoted to this topic, knowledge, especially for higher-dimensional systems (i.e. over $\mathbb{R}^n$, $n >> 1$), is still incomplete. Such phenomena are often used to justify the assertion that the majority of waves are not governed by the wave equation. Non-linear wave equations of interest, however, include ...

<u>Example 1.5</u> (Klein-Gordon)

A model of the following form is found in a variety of circumstances

$$\frac{\partial^2 z}{\partial t^2} - \frac{\partial^2 z}{\partial x^2} + V'(z) = 0 \qquad \ldots \qquad (1.15)$$

where $V'(\cdot)$ is a non-linear function of z occurring as the derivative of a potential energy $V(\cdot)$. Especially popular is the choice $V'(z) = \sin z$; predictably, this gives the "sin e - Gordon" equation. Such an equation has occurred in modelling dislocations in crystals, propagation of magnetisation waves in ferromagnetic materials, and Josephson junctions. With a cubic nonlinearity $(V'(z) = z^3)$ (1.15) has been used to make tentative suggestions about the nature of elementary particles. This should not be surprising in view of the close relation to the non-linear Schrödinger equation.

In all these cases we may be interested either in inserting control action to drive the system so as to meet certain objectives (e.g. attain a desired final state) or in taking output measurements so as to reconstruct the internal state. It is with this latter objective, namely state reconstruction, that this thesis will be mainly concerned. Many of the methods and theorems have analogies in the control case, but these are not explored here. Magnusson [1] investigates some applications of these methods to the control case. Thus we are here concerned to study a system model of the type (1.7) in the case where the state, $z'(\cdot)$, is infinite dimensional. The methods developed and the results obtained for infinite-dimensional state reconstruction in semi-linear systems, are new. In many cases these results are still new when restricted to a finite dimensional state-spaces. As indicated in Section 1.2 the output measurements will be taken to be given by a time invariant linear mapping acting on the system state. In systems governed by partial differential equations such a mapping could be, for example, the value (if well defined) of the system state at some point in its domain of definition or the result of integrating the system state over some sub-set of this domain.

1.4 Treatment

The preceding sections have already indicated the main features of the thesis : the special class of non-linear systems to be studied using fixed point results and strong assumptions on the linear part. The necessary linear theory is provided in Chapter 2. Chapter 3 concerns the basic results and Chapter 4 investigates various refinements. Originally Chapter 5 was going to be a rigorous study of the numerical analysis associated with application of the techniques of Chapters 3 and 4. This study, once embarked upon, proved to be both lengthy and intricate, demanding a variety of new material. A full presentation of this material was some way from the main aim of the thesis - this being, as previously stated, the application of fixed point results to semi-linear control theory. Thus, after some discussion, it was decided to restrict Chapter 5 to a largely formal account of these numerical aspects. This account tries to illustrate the main ideas and outline the requirements for a rigorous treatment. The purpose is to give an indication of this important area, without overburdening detail, and to show some promising directions for further research.

CHAPTER II : Linear Theory

Summary

This chapter presents a review of linear infinite dimensional control theory in a form suited to our later requirements. The basic material is drawn from Curtain-Pritchard [1] and Lions [2]. Additional material on analytic semigroups can be found in Hille-Phillips [1], on solvability and ill-posedness in Nashed [1], and on embedding theorems in Lions [3] and Adams [1]. None of the material is new, but its juxtaposition is slightly novel.

2.1 Linear Evolution Equations : semigroups

Consider the (finite dimensional) ordinary differential equation

$$\dot{z} = Az \quad ; \quad z(0) = z_0 \quad \ldots \tag{2.1}$$

where $z(t)$ lies in $\mathbb{R}^n$ for some n and A is an $n \times n$ matrix. The matrix exponential e^{At} is used to express the solution of (1) in the form

$$z(t) = e^{At} z_0 \quad \ldots \tag{2.2}$$

In the case of a linear partial differential equation we may have a representation similar to (2.1) where $z(t)$ now lies in some Banach space Z and wish to *define* a solution using an infinite dimensional analogue of the matrix exponential. This analogue is the semigroup.

More precisely we have

Definition 2.1

A strongly continuous semigroup is a map $S(\cdot)$ from $\mathbb{R}^+$ to $L(Z)$, satisfying

$$S(t+s) = S(t)S(s) \; ; \quad 0 \leq s \leq t \qquad \ldots \qquad (2.3)$$

$$S(0) = I \qquad \ldots \qquad (2.4)$$

$$||S(t)z_0 - z_0|| \to 0 \text{ as } t \to 0^+ \quad \forall\, z_0 \in Z \qquad \ldots \qquad (2.5)$$

Example 2.2

Let $A \in L(Z)$ and define

$$e^{At} = \sum_{n=0}^{\infty} \frac{(At)^n}{n!}$$

; this yields a strongly continuous semigroup.

Theorem 2.3

Let $S(t)$ be a strongly continuous semigroup on a Banach space Z, then

a) $||S(t)||$ is bounded on every finite subinterval of $[0,\infty[$

b) $\forall z \in Z$, $S(t)z$ is strongly continuous

c) if $w_0 = \inf_{t>0} \left(\frac{1}{t} \log ||S(t)|| \right)$ then $w_0 = \lim_{t\to\infty} \left(\frac{1}{t} \log||S(t)|| \right) < \infty$

d) $\forall w > w_0 \ \exists$ a constant M_w such that $\forall t \geq 0$

$||S(t)|| \leq M_w e^{wt}$

Pf. see Curtain-Pritchard, [1].

The connection between the semigroup and the solution of an abstract evolution equation is made by using the following

Definition 2.4

The infinitesimal generator A of a strongly continuous semigroup $S(t)$ on a Banach space Z is defined by

$$Az = \lim_{t \to 0+} \frac{1}{t} (S(t)-I)z \qquad \ldots \qquad (2.6)$$

whenever this limit exists; the domain of A, denoted $D(A)$ being the set of elements in Z for which the limit exists.

Theorem 2.5

Suppose $S(t)$ is a strongly continuous semigroup on a Banach space Z, with infinitesimal generator A, and thus

a) if $z_0 \in D(A)$ then $S(t)z_0 \in D(A) \ \forall t \geq 0$

b) $\frac{d}{dt}(S(t)z_0) = AS(t)z_0 = S(t)Az_0$ for $z_0 \in D(A)$, $t > 0$

c) $\frac{d^n}{dt^n}(S(t)z_0) = A^n S(t)z_0 = S(t)A^n z_0$, $z_0 \in D(A^n)$, $t > 0$

d) $S(t)z_0 - z_0 = \int_0^t S(s)Az_0 ds \qquad z_0 \in D(A)$

e) A is a closed linear operator, $\overline{D(A)} = Z$

f) $\bigcap_n D(A^n)$ is dense in Z

Pf. see Curtain-Pritchard [1].

Operators A which generate strongly continuous semigroups are characterised by the Hille-Yosida theorem. This theorem uses the

Definition 2.6

Let A be a closed densely defined linear operator. The set of complex numbers λ such that λ is not an eigenvalue and the range of $\lambda I - A$ is the whole space Z is called the resolvent set of A. For $\lambda \in \rho(A)$, $(\lambda I - A)^{-1}$ is denoted $R(\lambda;A)$ and is called the resolvent of A.

Theorem 2.7 (Hille-Yosida)

A closed linear operator, A, such that $\overline{D(A)} = Z$ for a Banach space Z generates a strongly continuous semigroup $S(t)$ iff $\exists$ real numbers M,w such that $\forall$ real $\lambda > w$, $\lambda \in \rho(A)$ the resolvent set of A and

$$||R(\lambda,A)^r|| \le \frac{M}{(\lambda-w)^r} \qquad r = 1,2,\ldots \qquad \ldots \qquad (2.7)$$

If this holds then

$$||S(t)|| \le Me^{wt}$$

Pf. see Hille-Phillips [1].

The conditions of this theorem are not always easy to check and other criteria have been developed; see Curtain and Pritchard [1], p.22 for example. As will be seen, in a variety of contexts (particularly in solvability and optimisation) adjoint operators naturally occur; in the following Z^* denotes the dual space of Z, $\langle\cdot,\cdot\rangle_{Z^*,Z}$ denotes the duality pairing.

Definition 2.8

Let A be a closed, densely defined linear operator with domain $D(A)$ in a Banach space Z. The adjoint operator A^* associated with A ("the adjoint of A") is a linear operator : $D(A^*) \to Z^*$ where

$$D(A^*) = \{z^* \in Z^* | \exists g^* \in Z^* : \langle g^*,z\rangle_{Z^*,Z} = \langle z^*,Az\rangle_{Z^*,Z} \quad \forall z \in D(A)\}$$

where we define $A^*z^* = g^*$.

As $D(A)$ is dense, A^*z^* is well defined. One use of the adjoint operator is in

Theorem 2.9

Let A be a closed densely defined linear operator on a Banach space Z, then A generates a semigroup $S(t)$ on Z satisfying $||S(t)|| \le e^{wt}$ $\forall t \ge 0$ iff $\forall \lambda > w$

$$||(\lambda I-A)z||_Z \ge (\lambda-w)||z||_Z \quad , \quad z \in D(A)$$

$$||(\lambda I-A^*)z^*||_{Z^*} \ge (\lambda-w)||z^*||_{Z^*} \, , \; z^* \in D(A^*)$$

Pf. see Curtain-Pritchard [1].

Corollary 2.10

In the case where Z is a Hilbert space if there exists a β such that

$$\beta||z||^2 \geq \text{Re}\,(\langle Az,z\rangle) \qquad z \in D(A)$$

$$\beta||z^*||^2 \geq \text{Re}\,(\langle A^* z^*,z^*\rangle) \qquad z^* \in D(A^*)$$

then A generates a semigroup $S(t)$ on Z .

Pf. Consider, for example, the condition

$$||(\lambda I - A)z||_Z \geq (\lambda - w)||z||_Z \qquad z \in D(A)$$

of Theorem 3.9. In the Hilbert space this is equivalent to ($\langle\cdot,\cdot\rangle$ now denotes inner product)

$$\langle \lambda z - Az \,,\, \lambda z - Az\rangle \geq (\lambda - w)^2\langle z,z\rangle$$

for $\lambda > w$, $z \in D(A)$; by expansion

$$2\lambda(w||z||^2 - \text{Re}\langle Az,z\rangle) + \langle Az,Az\rangle - w^2||z||^2 \geq 0$$

which will be satisfied for w,λ large enough if there exists a β as in the statement. Similarly for the other condition. ∎

Given a semigroup $S(t)$ it is often natural to consider its adjoint $S^*(t)$. What can be said about this collection of operators? A simple statement is...

Theorem 2.11

Let Z be a reflexive Banach space, $S(\cdot)$ a strongly continuous semigroup on Z with infinitesimal generator A. Then $S^*(\cdot)$ is a strongly continuous semigroup on Z^* with infinitesimal generator A^*.

Pf. see Curtain-Pritchard [1].

Example 2.12

Take $Az = -\frac{dz}{d\xi}$ $\quad Z = L^2([0,1];\mathbb{R})$

and $D(A) = \{z : z \in H^1([0,1];), z(0) = 0\}$.

Integration by parts gives that

$$A^*z = \frac{dz}{d\xi}$$

with $D(A^*) = \{z : z \in H^1([0,1];\mathbb{R}), z(1) = 0\}$

Additionally

$$\langle Az, z\rangle = -\tfrac{1}{2}(z(1))^2 \leq 0$$

$$\langle A^*z, z\rangle = -\tfrac{1}{2}(z(0))^2 \leq 0$$

Hence the conditions of Corollary 2.10 are satisfied with $\beta = 0$. Thus A generates a semigroup.

Example 2.13

Consider the system

$$\ddot{z} + \alpha\dot{z} + Az = 0 \quad ; \quad z(0) = z_0 \text{ , } \dot{z}(0) = z_1 \quad ; \alpha \geq 0$$

where A is a positive self-adjoint operator on a real Hilbert space H, with dense domain satisfying

$$<Az,z> \geq k||z||^2 \quad \forall z \in D(A) \text{ ; } \quad k > 0 \quad .$$

Proceeding formally we consider the first order system

$$\dot{w} = Qw \quad \text{where} \quad w = \begin{bmatrix} z \\ \dot{z} \end{bmatrix}$$

and

$$Q\begin{bmatrix} z \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -A & -\alpha \end{bmatrix} \begin{bmatrix} z \\ \dot{z} \end{bmatrix}$$

As A is self-adjoint and positive its square root $A^{\frac{1}{2}}$ is well defined and we may introduce a Hilbert space $\mathcal{H} = D(A^{\frac{1}{2}}) \times H$ with inner product

$$<w,\overline{w}>_{\mathcal{H}} = <A^{\frac{1}{2}}z,A^{\frac{1}{2}}\overline{z}>_H + <\dot{z},\dot{\overline{z}}>_H$$

and hence, for $w \in D(Q) = D(A) \times D(A^{\frac{1}{2}})$

$$\begin{aligned} <w,Qw>_{\mathcal{H}} &= <Az,\dot{z}> + <\dot{z},-Az-\alpha\dot{z}> \\ &= -\alpha||\dot{z}||_H^2 \quad . \end{aligned}$$

The adjoint of Q with respect to the Hilbert space H is given by

$$Q^* \begin{bmatrix} z \\ \dot{z} \end{bmatrix} = \begin{bmatrix} 0 & -I \\ A & -\alpha \end{bmatrix} \begin{bmatrix} z \\ \dot{z} \end{bmatrix}$$

$$D(Q^*) = D(Q) .$$

Therefore $\langle w, Q^* w \rangle = -\alpha ||\dot{z}||_H^2$; hence, by Corollary 2.10 we have that Q generates a semigroup on H (in fact, a strongly continuous semigroup).

Example 2.14 (Wave equation)

This is a special case of the preceding.

Take $z_{tt} = z_{xx}$

with $z(0,t) = z(1,t) = 0$. Let $H = L^2([0,1];\mathbb{R})$ and $Az = -z_{xx}$ (in the formulation of the preceding)

$$D(A) = H^2(0,1) \cap H_0^1(0,1)$$

then $A^* = A$ and (ξ is a dummy spatial integration variable)

$$\langle z, Az \rangle_H = \int_0^1 z_x^2(\xi) d\xi \geq \pi^2 \int_0^1 z^2(\xi) d\xi = \pi^2 ||z||_H^2$$

using integration by parts and standard embeddings.
In the formulation of Ex. 2.13 we have

$$Q = \begin{bmatrix} 0 & I \\ -A & 0 \end{bmatrix} \begin{bmatrix} z \\ \dot{z} \end{bmatrix}$$

$$D(Q) = D(A) \times D(A^{\frac{1}{2}}) = H^2(0,1) \cap H_0^1(0,1) \times H_0^1(0,1)$$

and can conclude that Q generates a strongly continuous semigroup $S(t)$ on H. If we separate the componenets of the domain space as w_1(for z) and w_2(for $\dot{z}$) then

$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in H_0^1(0,1) \times L^2(0,1)$ and we have the following explicit expression for the semigroup action

$$S(t)\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \Sigma 2[<w_1,\phi_n>\cos n\pi t + \frac{1}{n\pi}<w_2,\phi_n>\sin n\pi t]\phi_n \\ \Sigma 2[-n\pi<w_1,\phi_n>\sin n\pi t + <w_2,\phi>\cos n\pi t]\phi_n \end{bmatrix}$$

where $\phi_n = \sin n\pi\xi$.

Example 2.15 (Heat equation)

The heat equation

$$z_t = z_{xx}$$

$$z(0,t) = z(1,t) = 0$$

$$z(x,0) = z_0$$

is a prototype for a large class of parabolic equations.
Taking $H = L^2(0,1)$ one has $Az = z_{xx}$ defined on the domain $D(A) = H^2(0,1) \cap H_0^1(0,1)$. In this case the semigroup has the explicit expression

$$(S(t)z)(\xi) = \Sigma 2e^{-n^2\pi^2 t}\sin n\pi\xi \int_0^1 \sin n\pi y \, z(y)dy .$$

The fact that the A in question generates a strongly continuous semigroup can be deduced from Hille-Yosida Theorem - see Curtain-Pritchard [1]. Semigroups generated by elliptic operators possess additional smoothing properties, which are briefly described in the appendix on analytic semigroups.

Notion of a solution : semigroup

Using Theorem 2.5 it is clear that if z_0 lies in $D(A)$, where A generates a strongly continuous semigroup $S(t)$, we can define a solution of $\dot{z} = Az$ as $z(t) = S(t)z_0$. This solution is continuous on $[0,\infty[$, differentiable on $]0,\infty[$, and unique. It is often referred to as the *strong* solution. If, however, $z_0 \notin D(A)$ we may still wish to have some notion of solution - in this case one *defines* the mild solution as $S(t)z_0$.

Example 2.16

In regard of example 2.14, the mild solution is given by

$$\begin{bmatrix} z \\ z_t \end{bmatrix} = S(t) \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \quad \text{for} \quad z_0 \in H_0^1(0,1) ,$$

$z_1 \in L^2[0,1]$. In case that $z_0 \in H^2(0,1) \cap H_0^1(0,1)$, $z_1 \in H_0^1(0,1)$ then this solution is in fact a strong solution.

Suppose now one considers the inhomogeneous equation

$$\dot{z} = Az + f \; ; \quad z(0) = z_0 \; , \qquad \ldots \qquad (2.8)$$

then, by analogy with finite dimension (variation of constants formula), one has, Definition 2.17 (Mild solution)

If $f \in L^p(0,t_1;Z)$ $p \geq 1$ then

$$Z(t) = S(t)z_0 + \int_0^t S(t-s)f(s)ds \quad \ldots \quad (2.9)$$

is a mild solution of (2.8) on $[0,t_1]$.

One can now show

Lemma 2.18

$z(t)$ defined by (2.9) is strongly continuous on $[0,t_1]$.

Pf. see Curtain-Pritchard [1].

This lemma is important because it tells us for a useful class of functions f where the resulting state trajectory will lie. There are other notions of solution relevant to (2.9), in particular, that investigated in Lions-Magenes [1]. Thus far we have attempted to present a formulation which applies to both the hyperbolic (Examples 2.13, 2.14) and the parabolic (Example 2.15) cases. This generality is not always desirable since the two cases possess some fundamental differences. In particular, some hyperbolic equations (Example 2.14) will give rise to a semigroup which is in fact a group; whereas, since a parabolic equation smooths the initial data, one cannot invert the semigroup action. These statements will be made more precise in the following. Lions [2] presents a framework for linear parabolic equations as follows.

Notion of a solution : weak

Let V,H be Hilbert spaces such that V is continuously embedded and dense in H ; V^* is the dual of V and H is identified with its dual H^*. One writes $V \subset H \subset V^*$ and considers a family of operators $A(t):V \to V^*$ (hence defining a map $A(\cdot) \in L(L^2(0,t_1;V);L^2(0,t_1;V^*))$ by $f \to A(t)f(t)$ for $f \in L^2(0,t_1;V))$. The objective is to study the evolution equation

$$\frac{dz}{dt} + A(t)z = f \qquad \ldots \qquad (2.10)$$

with initial conditon

$$z(0) = z_0 \qquad \ldots \qquad (2.11)$$

for appropriately chosen f, z_0. To do this one introduces

Definition 2.19

$$W(0,t_1) = \{f | f \in L^2(0,t_1;V), \frac{df}{dt} \in L^2(0,t_1;V^*)\}$$

with the norm

$$||f||_{W(0,t_1)} = \left(\int_0^{t_1} ||f(s)||_V^2 ds + \int_0^{t_1} ||\frac{df}{dt}||_{V^*}^2 ds\right)^{\frac{1}{2}} .$$

Using the fact that V is Hilbert, $W(0,t_1)$ normed as above can be given a Hilbert space structure. The following regularity result is important for what follows.

Theorem 2.20

Any $f \in W(0,t_1)$ is, after possible modification on a set of measure zero, a member of $C(0,t_1;H)$,

Pf. see Lions-Magenes, [1], Volume 1.

Take, in (2.11), $z_0 \in H$, $f \in L^2(0,t_1;V^*)$ in (2.10) and $A(t)$ satisfying the "coercivity condition" $\exists\ \lambda$ such that

$$\langle A(t)\phi,\phi\rangle_{V^*,V} + \lambda||\phi||_H \geq \alpha||\phi||_V \quad \alpha > 0\ , \quad \forall\phi \in V\ , \ t \in\ ]0,t_1[\ .$$

<u>Theorem 2.21</u>

With f,z_0 , $A(t)$ as above, the problem (2.10), (2.11) has an unique solution in $W(0,t_1)$. Moreover the induced map

$$f,z_0 \rightarrow z \text{ is continuous from}$$

$$L^2(0,t_1;V^*) \times H \rightarrow W(0,t_1)$$

Pf. see Lions, [2].

The coercivity condition used above ensures that when $A(t)$ is time invariant (i.e. $A(t) = A\ \ \forall t \in (0,t_1))$, $-A$ generates a strongly continuous semigroup. Hence, by Theorem 2.20, Theorem 2.21 is consistent with the semigroup approach (where one obtains $z \in C(0,t_1;H)$. Note that a different class of perturbations is used, i.e. $f \in L^2(0,t_1;V^*)$ rather than $L^p(0,t_1;H)$.

2.2 Solvability and least squares problems : generalised inverses

Consider the operator equation

$$Tx = y \qquad \ldots \qquad (2.12)$$

where T is a linear mapping from a Hilbert space X into a Hilbert space Y. Traditionally one considers that (2.12) has a solution if and only if $y \in$ range (T). In a variety of circumstances (particularly in optimisation problems) this traditional notion may not be the most appropriate. The generalized (or sometimes pseudo) inverse offers one way in which the notion of a solution may be extended.

Definition 2.22

T is said to be a closed operator iff $G(T)$, the graph of T, is closed in $X \times Y$ (i.e. $x_n \in X$, $x_n \to x_0 \Rightarrow Tx_n \to y_0$: $x_0 \in X$ and $Tx_0 = y_0$).

Clearly, any $T \in L(X,Y)$ is a closed operator. Note that the range of T is not necessarily closed in Y. Now and for the rest of this section we identify X^* with X and Y^* with Y; T^* will denote the adjoint map; then we have that

$$\overline{R(T)} = N(T^*)^{\perp} \qquad R(T)^{\perp} = N(T^*)$$
$$\ldots \qquad (2.13)$$
$$\overline{R(T^*)} = N(T)^{\perp} \qquad R(T^*)^{\perp} = N(T)$$

Clearly, $R(T)$ is closed in Y iff $R(T^*)$ is closed in X.

$$T|_{N(T)^{\perp}} : N(T)^{\perp} \to R(T) \quad \text{is} \quad 1 - 1 \; ; \quad \text{thus}$$

$(T|_{N(T)^{\perp}})^{-1}$ exists and is such that

$$(T|_{N(T)^{\perp}})^{-1} : R(T) \to N(T)^{\perp}$$

It is now reasonable to define the generalized inverse $T^\dagger$ of T as

Definition 2.23

$T^\dagger$ is the linear extension of $(T|_{N(T)^\perp})^{-1}$ so that

$D(T^\dagger) = R(T) + R(T)^\perp$

$N(T^\dagger) = R(T)^\perp$.

If $R(T)$ is closed then $Y = R(T) + R(T)^\perp$ so that $T^\dagger$ is a bounded operator; if $R(T)$ is not closed then $R(T^\dagger)$ is $\overline{R(T^*)}$.

As might be expected in a Hilbert space, the generalized inverse has a natural interpretation in terms of projections. Let P denote the orthogonal projection of X onto $N(T)^\perp$ and Q the orthogonal projection of Y onto $N(T^*)^\perp$. It can be shown that

$$P = T^\dagger T \quad \ldots \quad (2.14)$$

$$Q = T\,T^\dagger \quad \ldots \quad (2.15)$$

Thus the projection onto $N(T^*)$ may be written as $I - TT^\dagger$ and that onto $N(T)$ as $I - T^\dagger T$. Further properties of generalized inverses will be found in Appendix 3.

Least squares problems

One of the most useful features of generalized inverses is their relationship to least squares problems. Consider $Tx = y$, $T \in L(X,Y)$

once more; suppose we now no longer wish to find a solution x which satisfies $Tx = y$ exactly, instead we are content to find an x which minimises $||Tx - y||_Y$.

Definition 2.24

A vector x^{LS} is a least squares solution if

$$||Tx^{LS} - y||_Y = \inf\{||Tx - y||_Y : x \in X\} .$$

All such x^{LS} satisfy $T^*T x^{LS} = T^*y$; the so-called normal equations.

Definition 2.25

A vector $\hat{x}$ will be called a least squares solution of minimum norm if $\hat{x}$ is a least squares solution and

$$||\hat{x}||_X \leq ||x^{LS}||_X \text{ for all least squares solutions } x^{LS} .$$

The set of least squares solutions may be empty. If $R(T)$ is closed, however, the set of all x^{LS} is non-empty closed and convex; and a closed convex set in a Hilbert spaces possesses an unique element of minimum norm. The relationship between the generalized inverse and the least squares solution of minimum norm is expressed by

Theorem 2.26

Let T be bounded, $R(T)$ closed, then

$$\hat{x} = T^{\dagger} y$$

Pf. see Beutler [1]

and

Theorem 2.27

If T is bounded, but $R(T)$ not necessarily closed then if $y \in D(T^{\dagger}) = R(T) + R(T^{\dagger})$ we have $\hat{x} = T^{\dagger}y$.

Pf. see Beutler [1] .

Of course, many problems can be formulated in a linear least squares context. An operator which gives the solution to such a problem may offer no inherent computational advantages. It may be possible to use the algebraic identities satisfied by the pseudo inverse to solve the linear least squares problem; but it is more likely that, especially in the infinite dimensional case, one would resort to some more traditional minimisation procedure. It is true, however, that existence of such an operator offers some analytic advantages.

Solvability

The solvability of (2.12) has been considered, in the mathematical literature, since the time of Hausdorff and Fredholm. Take $T \in L(X,Y)$ with $X,\tilde{Y}$ Hilbert and make the . . .

Definition 2.28

T is normally solvable iff $R(T)$ is closed.

This is equivalent to (using (2.13)).

Definition 2.29

The equation $Tx = y$ is consistent iff y is orthogonal to any solution u of $T^*u = 0$.

For bounded linear T such as are considered here one obtains

Theorem 2.30

The following are equivalent

a) $R(T)$ is closed;

b) $\gamma(T) = \inf.\{\frac{||Tx||}{||x||} : 0 \neq x \in N(T)^{\perp}\} > 0$;

c) $\inf\{||Tx-y|| : x \in X\}$ is attained $\forall y \in Y$;

d) the restriction of T to $N(T)^{\perp}$ has a bounded inverse;

e) the quotient space $X/N(T)$ is isomorphic with $R(T)$;

f) T has a bounded generalized inverse.

Pf. see Nashed [1].

In the case of linear operators $T:X \rightarrow Y$ which are unbounded but have closed graphs, a theorem analogous to the above holds, subject to certain modifications. For example, in condition b) x must be restricted to lie in $D(T) \cap N(T)^{\perp}$. Note also that normal solvability for any one of T, T^*, TT^*, or T^*T implies the same for all the others.

Example 2.31

Examples of normally solvable T are given by

a) all operators which are bounded below (i.e. $||Tx|| \geq m||x||$ for some $m > 0$;

b) all operators of the form $T = T_1 - \lambda T_2$, $\lambda \neq 0$, where T_2 is completely continuous and T_1 has a bounded inverse;

c) all operators of the form $T = T_1 + T_2$ where $R(T_1)$ is closed and $R(T_2)$ is finite dimensional.

These ideas give rise to some simple, but sometimes ignored, conclusions of relevance to the construction of algorithms.

Algorithmic implications

Several different versions of well (or ill) posedness exist in the literature. That appropriate to the present setting is . . . .

Definition 2.32

The equation $Tx = y$ ($T \in L(X,Y)$; X,Y being Hilbert) is said to be well-posed, relative to the spaces X and Y if, for each $y \in Y$ the unique "solution" $T^{\dagger}y$ depends continuously on y ; otherwise the equation is said to be ill-posed.

Then one can show

Theorem 2.33

Let T be as above. The following are equivalent

a) the operator equation $Tx = y$ is well-posed relative to the spaces X and Y ;

b). T has closed range in Y .

Pf. see Nashed [1].

In the rest of this section, both for simplicity of presentation, and because it is the only case which will be studied in detail in the sequel, we shall assume that T is injective (i.e. $N(T) = \{0\}$) . Suppose now that y is in the range of T i.e. $\exists \; x \in X : Tx = y$. Suppose also that in practice only an approximate $y_\varepsilon : ||y - y_\varepsilon|| \leq \varepsilon$ is available, then the solution (if it exists) x_ε such that $Tx_\varepsilon = y_\varepsilon$ need not be close to x for ε close to zero. Boundedness of the inverse map would prevent such pathological behaviour. In any case, even if y is known exactly any discretised version of the problem will still be badly ill-conditioned. Thus one needs to develop, at the very least, some method of producing approximate problems which always have a solution and additionally, if possible, a solution which depends continuously on the given data, y . A variety of standard remedies exist - we choose to present

<u>Definition 2.34</u>

The augmented problem $P_{\varepsilon,X}$, $\varepsilon > 0$, is defined as "find x of minimum norm in X which minimizes

$$||Tx - y||_Y^2 + \varepsilon^2 ||x||_X^2 \text{"}$$

From the section on least squares problems, $P_{\varepsilon,X}$ corresponds to finding the generalized inverse for the augmented operator

$$\tilde{T}_{\varepsilon,X} : X \to X \times Y : x \to (x, Tx)$$

where $X \times Y$ is given the norm $||\cdot||_Y^2 + \varepsilon^2||\cdot||_X^2$.

The generalized inverse of this operator will be denoted by $\tilde{T}^{\dagger}_{\varepsilon,X}$. Note that as T is a closed operator $\tilde{T}_{\varepsilon,X}$ has closed range and therefore a bounded generalized inverse. One can consider that the problem $P_{\varepsilon,X}$ concerns a normal equation

$$(T^*T + \varepsilon I)x = T^*y$$

(for $\varepsilon > 0$, $T^*T + \varepsilon I$ will be invertible). It can be shown (see for example Lions-Stampacchia [1]) that if $y \in$ range T then, as $\varepsilon \to 0^+$, $\tilde{T}^{\dagger}_{\varepsilon,X}(0,y) \to x$ where $Tx = y$. We still do not obtain, in the limit, any continuity with respect to the data y . Such a property is obtained by using a device due to Tikhonov [1], which is now presented.

Consider a Hilbert space Z which is compactly embedded in X . Then by analogy with Definition 2.34 we have

<u>Definition 2.35</u>

The augmented problem $P_{\varepsilon,Z}$ is defined as "find x_ε of minimum norm in Z which minimizes $||Tx - y||_Y^2 + \varepsilon^2||x||_Z^2$ "

Denote by $\tilde{T}_{\varepsilon,Z}$ the associated member of $L(Z,Z\times Y)$ and by $\tilde{T}^{\dagger}_{\varepsilon,Z}$ its generalized inverse. From the results of Tikhonov [1] one obtains the following continuity property.

Theorem 2.36

Let y be in $T(Z)$ (i.e. $\exists\ x_* \in Z : Tx_* = y$). Let $y_\varepsilon \to y$ in Y as $\varepsilon \to 0^+$ and define $x_\varepsilon = \tilde{T}^{\dagger}_{\varepsilon,Z}(0,y_\varepsilon)$. Then $x_\varepsilon \to x_*$ in X .

Pf. see Tikhonov [1].

The above procedure describes one possible regularisation method for ill-posed problems; it is a method which, as will be seen, is peculiarly appropriate to the present requirements. For further information on regularisation see Ribière [1].

2.3 Linear infinite dimensional control theory : formulation

Knowledge of linear evolution equations in function spaces (e.g. definition and representation of solutions; allowable classes of perturbations) and of linear least squares problems (posed in Hilbert spaces) is now applied to the study of some problems in control theory. As indicated in Chapter I, the scope of the treatment will rapidly be restricted to problems of state reconstruction. The control case gives rise to slightly different but naturally related considerations; in the linear theory this relationship is, in effect, that between a linear map and its adjoint; and, in particular, between the range of the map and the kernel of its adjoint. The control and observation problems will both be stated first in the semigroup case and then for the Lions formulation of 2.2. The latter formulation, as stated in 2.2, only applies to parabolic problems; though there are extensions, see Lions [1], to second order hyperbolic equations such as Example 2.14. We shall not investigate these extensions here and will for the rest of this thesis be concerned with three classes of linear problems viz.

a. second order hyperbolic equations, in the semigroup formulation;

b. parabolic equations, in the case where both the semigroup formulation and the Lions formulation apply;

c. finite dimensional equations where, of course, many of these problems of formulation disappear.

Control and observation : semigroup

Consider

$$\dot{z} = Az + Bu \quad \ldots \quad (2.16)$$

$$z(0) = z_0 \ , \quad \ldots \quad (2.17)$$

on the interval $[0,t_1]$, where A generates a strongly continuous semigroup $S(t)$ on a Banach space Z, $z_0 \in Z$, and B is a bounded operator from a space of controls U into Z. We make the following.

Definition 2.37

The system (2.16), (2.17) is said to be controllable on $[0,t_1]$ iff given any two points $z_0, z_1 \in Z$ there exists a control $u(\cdot) \in L^p(0,t_1;U)$ $(p \geq 1)$ such that $z(0) = z_0$ and $z(t_1) = z_1$.

Define the operator

$$G_c : L^p(0,t_1;U) \to Z : u \to \int_0^{t_1} S(t_1-s)B\, u(s)ds$$

and introduce the following definition (the need for this definition will be indicated in the sequel).

Definition 2.38

The system (2.16), (2.17) is said to be approximately controllable on $[0,t_1]$ iff $\overline{\text{range}\ (G_c)} = Z$.

Hence (2.16), (2.17) is approximately controllable in time t_1 if, for any $z_1 \in Z$, and any $\varepsilon > 0$, $\exists\, u(\cdot) \in L^p(0,t_1;U)$ such that $||z(t_1) - z_1|| \leq \varepsilon$.

Consider now the observation problem given by

$$\dot{z} = Az \qquad \ldots \qquad (2.18)$$

$$y = Cz \qquad \ldots \qquad (2.19)$$

again over the time interval $[0,t_1]$; A, as before, generates a strongly continuous semigroup $S(t)$ on a Banach space Z and C is a bounded operator from Z, the space of states, into Y, a space of outputs. Hereinafter we shall assume Z,Y (and U) to be reflexive Banach spaces. Given the observed outputs $y(\cdot)$ one wishes to reconstruct the appropriate initial state z_0. Define the operator

$$H_0 : Z \to L^q(0,t_1;Y) : z_0 \to CS(t)z_0 \ (q > 1) \quad \text{and}$$

introduce the following ...

<u>Definition 2.39</u>

The system (2.18), (2.19) is said to be initially observable on $[0,t_1]$ iff $N(H_0) = \{0\}$.

In practice, however, we may wish for the existence of a continuous reconstruction operator

$$H_1 : \text{range } (H_0) \to Z : H_1 H_0 = I \ .$$

When range (H_0) has the induced topology from $L^q(0,t_1;Y)$, the existence of such an operator is ensured by

Definition 2.40

The system (2.18), (2.19) is said to be continuously initially observable on $[0,t_1]$ iff $\exists$ $\gamma \in \mathbb{R}$, $\gamma > 0$ such that

$$\gamma ||H_0 z||_{L^q(0,t_1;Y)} \geq ||z||_Z \qquad \forall z \in Z .$$

It is clear that to each system (2.18), (2.19) one can associate a "dual" controlled system

$$\dot{z} = A^* z + C^* u$$

where $u(\cdot) \in L^p(0,t_1;Y^*)$ $p : \frac{1}{p} + \frac{1}{q} = 1$ to which one can apply the considerations of the control section. The following theorem results.

Theorem 2.41

i) (2.18), (2.19) is initially observable on $0,t_1$ iff the dual controlled system is approximately controllable on $[0,t_1]$.

ii) (2.18), (2.19) is continuously initially observable on $[0,t_1]$ iff the dual controlled system is exactly controllable on $[0,t_1]$.

Pf. see Curtain-Pritchard, [1].

This duality expresses the fact that the adjoint of G_c is an operator of type H_0 (but for the dual system). The surjectivity of G_c

corresponds to the injectivity of H_0 ; and G_c is a closed operator iff H_0 is. This duality, and the obvious relationships to section 2.2, will be explored in the sequel. For future reference, we present four examples.

Example 2.42

Let Z be a real separable Hilbert space and consider an operator A defined by $(R : 1 \le k \le r_n < \infty$ and ϕ_{nk} orthonormal set)

$$Az = \sum_{n=1}^{\infty} \lambda_n \sum_{k=1}^{r_n} \phi_{nk} \langle \phi_{nk}, z \rangle \quad \ldots \quad (2.20)$$

the conditions of Example 2.12, Curtain-Pritchard [1] must be satisfied for A to generate a semigroup.
This corresponds to A possessing eigenvalues $\lambda_n (\lambda_1 > \lambda_2 > \ldots)$ of (finite) multiplicity r_n . Let the control system have the form

$$\dot{z} = Az + \sum_{j=1}^{m} b_j u_j \quad \ldots \quad (2.21)$$

where $b_1, \ldots, b_m \in Z$ and $u_j \in L^p(0,t_1)$, $1 < p < \infty$.
Then $B(u_1, \ldots, u_m) = \sum_{j=1}^{m} b_j u_j$ and

$$B^* z = (\langle b_1, z \rangle_Z , \ldots , \langle b_m, z \rangle_Z) \ .$$

Hence by duality the mild solution of (2.21) is approximately controllable on $[0,t_1]$ iff

$$\sum_{n=1}^{\infty} e^{\lambda_n t} \sum_{k=1}^{r_n} \langle b_j, \phi_{nk} \rangle \langle z^*, \phi_{nk} \rangle = 0 \quad \ldots \quad (2.22)$$

$$(j = 1, \ldots, m \ ; \ t \in [0,t_1])$$

implies $z^* = 0$.

Using (2.22) one can deduce (see Curtain-Pritchard [1]) the following result "(2.21) is approximately controllable on $[0,t_1]$ iff rank $B_\ell = r_\ell$ " where B_ℓ is the matrix

$$B_\ell = \begin{bmatrix} \langle b_1,\phi_{\ell 1}\rangle_Z, & \cdots\cdots & ,\langle b_m,\phi_{\ell 1}\rangle_Z \\ & & \\ \langle b_1,\phi_{\ell r_\ell}\rangle_Z, & \cdots\cdots & ,\langle b_m,\phi_{\ell r_\ell}\rangle_Z \end{bmatrix}$$

Hence the number of controls required is at least as great as the highest multiplicity of the eigenvalues.

Example 2.43 (cf Example 2.15)

$$z_t = z_{xx} + b_1(x)u(t) \qquad\qquad u(t) \in \mathbb{R}$$

$$z(0,t) = z(1,t) = 0$$

has the dual

$$z_t = z_{xx}$$

$$z(0,t) = z(1,t) = 0$$

$$y(t) = \int_0^1 b_1(x)z(x,t)dx \quad .$$

The operator A is self-adjoint with compact resolvent and the eigenvalues and eigenvectors are $\lambda_n = -n^2\pi^2$ and $\phi_{n1} = \sin n\pi x$ and $t_n = 1$ for all n. Thus the controlled system will be approximately controllable using Example 2.42, with only one control if

$$\int_0^1 b_1(x) \sin n\pi x \ \ dx \neq 0 \qquad \forall n$$

By duality the same condition ensures initial observability for the dual system.

<u>Example 2.44</u>

$$z_t = z_{xx} + u$$

$$z(0,t) = z(1,t) = 0 \ .$$

Take $u(\cdot) \in L^2(0,t_1;L^2(0,1))$ then, using the expressions for the semi-group in Example 2.15, it can be shown that the system is exactly controllable to $H_0^1(0,1)$, but not $L^2(0,1)$. See the demonstration in Curtain-Pritchard [1], p.59, for example. By duality the observation problem with "complete information", i.e. output considered in $L^2(0,t_1;L^2(0,1))$ is only initially observable.

<u>Example 2.45</u> (cf Example 2.14)

$$z_{tt} = z_{xx} + u(x,t)$$

$$z(0,t) = z(1,t) = 0 \ ; \ z(x,0) = z_t(x,0) = 0 \ ,$$

Using the analysis of 2.14 we may put this problem within our present framework, by augmenting the state and creating $B = \begin{bmatrix} 0 \\ I \end{bmatrix}$ with $u \in L^2(0,t_1;L^2(0,1))$ we get that the system is exactly controllable on $[0,t_1]$ for any $t_1 > 0$. Hence the dual problem (i.e. complete observation of z_t) is continuously initially observable.

Control and observation : weak solution

Here the control and observation problems are stated in the form used by Lions [2]. The weak or variational formulation is naturally suited to the study of linear least squares problems; and, indeed, that is how control and observation are formulated. Recalling the material on weak solutions in 2.2, one defines $\mathcal{U}$, a Hilbert space of controls (for example previously we have used $\mathcal{U} = L^2(0,t_1;U)$ for U some Hilbert space) and B a bounded linear operator such that

$$B \in \mathcal{L}(\mathcal{U}, L^2(0,t_1;V^*)) .$$

As before, let f and z_0 be given : $f \in L^2(0,t_1;V^*)$ and $z_0 \in H$. Now, supposing that the family $A(t)$ satisfies the required coercivity conditions, see Theorem 2.21, one obtains for $u \in \mathcal{U}$ $z(u)$ satisfying

$$\frac{dz}{dt} + A(t)z = f + Bu \quad \ldots \quad (2.23)$$

$$z\big|_{t=0} = z_0 \quad \ldots \quad (2.24)$$

$$z \in L^2(0,t_1;V) \quad \ldots \quad (2.25)$$

Consider now

$$\underset{\sim}{C} \in \mathcal{L}(W(0,t_1);\mathcal{Y}) \quad \ldots \quad (2.26)$$

where $\mathcal{Y}$ is a Hilbert space of outputs (for example, $\mathcal{Y} = L^2(0,t_1;Y)$ where Y is Hilbert), $y_d \in \mathcal{Y}$ and

$$M \in \mathcal{L}(\mathcal{U};\mathcal{U}) \quad \ldots \quad (2.27)$$

such that

$$\langle Mu,u \rangle_{\mathcal{U}} \geq \nu ||u||^2_{\mathcal{U}} \qquad \nu \in \mathbb{R}, \ \nu > 0 \qquad \ldots \qquad (2.28)$$

then standard results give the following.

Theorem 2.46

With the above assumptions there exists an unique u minimising $||Cy(u) - y_d||^2_{\mathcal{Y}} + \langle Mu,u \rangle_{\mathcal{U}}$.

Pf. see Lions [2].

In [2] Lions indicates how to formulate the adjoint system for this problem. The adjoint system gives an explicit expression for the optimally controlled system; in numerical work the adjoint is used to find the gradient for use in a variety of algorithms. The treatment of Lions has other refinements, for example the consideration of arbitrary closed convex subsets of the space $\mathcal{U}$ as admissible controls, but the above will suffice for our purposes.

The observation problem is now formulated by viewing the initial condition as a control. That is, the space of admissible controls $\mathcal{U}$ is taken equal to H . As before one has $C \in \mathcal{L}(W(0,t_1);\mathcal{Y})$, $y_d \in \mathcal{Y}$, and $M \in \mathcal{L}(H;H)$ with M coercive i.e. satisfying (2.28), and then one obtains

Theorem 2.47

The problem : find $z(u)$ satisfying

$$\frac{dz}{dt} + A(t)z = f \qquad \ldots \qquad (2.29)$$

$$z\big|_{t=0} = u \quad \ldots \quad (2.30)$$

which minimizes

$$J(u) = ||Cz(u) - y_d||_{\mathcal{Y}}^2 + \langle Mu,u\rangle_H \quad \ldots \quad (2.31)$$

:has an unique solution. Moreover this can be characterised by an optimality system (i.e. equations for evolution of the state and its adjoint).

Pf. Lions [2].

Equations (2.29), (2.30), (2.31) indicate one approach to the problem of initial state reconstruction. It is possible to formulate more general reconstruction problems - e.g. versions which introduce some notion of model error - some comments on these directions will be made in the next paragraph.

Relationship to least squares formulation

Consider now that the hypotheses of the Lions formulation hold; then the problems posed by Lions (e.g. Theorems 2.46, 2.47) naturally correspond to the explicit construction of certain pseudo-inverses. For example, in Theorem 2.47, consider the map T defined from H to $\mathcal{Y}$ by $T:u \to Cz(u)$ (this map is "identical" with H_0 of Definition 2.39 with Z taken equal to H and $L^2(0,t_1:Y)$ as $\mathcal{Y}$). If the system is initially observable but not continuously initially observable T does not have closed range, although it is injective. Suppose however (without any real loss of generality) that $M \equiv \varepsilon^2 I_H$ (I_H denoting the identity map on H, $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$). Then one has, using the notation and results following Definition 2.34, ...

Theorem 2.48

The problem $P_{\varepsilon,H}$ is that described in Theorem 2.47 and thus for all $y_d \in \mathcal{Y}$ has solution given by $\tilde{T}^{\dagger}_{\varepsilon,H}(0,y_d)$. Moreover if $y \in$ range T then as $\varepsilon \to 0^+$ $\tilde{T}^{\dagger}_{\varepsilon,H}(0,y) \to x_*$ in H such that $Tx_* = y$.

Pf. cf Lions-Stampacchia [1].

Suppose now that one wishes to include some model error in the formulation. For instance, the setting of (2.29), (2.30) and (2.31) yields two possible ways of proceeding:

a. by regarding the model error as a "control" one obtains the problem:

find z(u,v) satisfying

$$\frac{dz}{dt} + A(t)z = f + v \quad \ldots \quad (2.32)$$

$$z\Big|_{t=0} = u \quad \ldots \quad (2.33)$$

which minimizes

$$J(u,v) = ||Cz(u,v) - y_d||^2_{\mathcal{Y}} + \varepsilon^2(\langle u,u\rangle_H + \langle v,v\rangle_{L^2(0,t_1;V^*)}) \quad \ldots \quad (2.34)$$

b. by regarding the state trajectory as available for choice one obtains the problem:

define $T : W(0,t_1) \times H \to L^2(0,t_1;V^*) \times \mathcal{Y}$

by $T : (z,u) \to (\frac{dz}{dt} + A(t)z \ , \ Cz)$

where $z\Big|_{t=0} = u$ $\quad \ldots \quad (2.35)$

Find (z,u) which minimizes

$$J(z,u) = ||\frac{dz}{dt} + A(t)z - f||_{L^2(0,t_1;V^*)} + ||Cz - y_d||_y^2 + \varepsilon^2(\langle z,z\rangle_{W(0,t_1)} + \langle u,u\rangle_H) \quad \ldots \quad (2.36)$$

Note that if y_d is such that $y_d = Cz$, where z is a solution of (2.29), (2.30), then as $\varepsilon \to 0^+$ both a. and b. give the "same" solution. Note also that both a. and b. can be regarded as concerning the construction of pseudo-inverse operators. In the case of (2.36) the construction in question is that of (using the notation of Definition 2.34) $\tilde{T}^{\dagger}_{\varepsilon,W(0,t_1)\times H}(0,0,f,y_d)$. This operator will be used in Chapter 4.

As to interpretation of (2.34), (2.36); if we let $\varepsilon \to 0$ in (2.34) we see that we obtain an "output error only" formulation - although the state trajectory (and thus the output trajectory) now depends not only on the initial state, u, but also on a perturbation, v, of the differential equation. So (2.34) could be regarded as a "regularised output-error formulation". By contrast in (2.36), letting $\varepsilon \to 0$, one obtains a problem containing both output error and "model error" weightings. The advantages of schemes involving model error weightings are well-known. In a closely related area Jazwinski, [1], discusses difficulties with the standard numerical interpretation of the Kalman filter when model error vanishes.

Problem a. falls immediately within the framework of Lions [2]; it is not explicitly studied in Lions [2] as only the simpler problems involving either u or v (but not both) are considered there. Not so for problem b. since the Lions treatment does not admit of z as an independent variable. One can, using the notation of a. replace (2.36) by

$$J(u,v) = ||Cz - y_d||_y^2 + ||v||^2_{L^2(0,t_1;V^*)} +$$

$$\varepsilon^2(\langle v,v\rangle_{L^2(0,t_1;V^*)} + \langle u,u\rangle_H)$$

a formulation, having similar properties to (2.36), and which falls within the Lions treatment. For later use, however, we prefer an "explicit extraction of z".

2.4 Linear least squares analysis : further refinements

The preceding section has shown how problems of control and observation can be cast in a linear least squares framework. The solution of these problems can then be represented by some pseudo-inverse. In the case (common in infinite dimensions) that the operator, for whose pseudo-inverse one searches, does not have closed range, then we construct a regularisation (using in effect the closed graph theorem and the graph norm) which has a bounded pseudo-inverse. The regularised problem is not an artificial one - as can be seen in 2.3, the regularised problem is often an entirely natural control or estimation problem, well-known in its own right. In this section we shall look at further properties of the regularisation. To do so we shall require that the hypotheses of the Lions formulation hold; and, in addition, that the family $A(t)$ be time invariant i.e. $A(t) = A$, $\forall t \geq 0$. Thus it will also be possible to make use of the semigroup formulation and notations. Latterly some other aspects of infinite dimensional problems are briefly described.

More regularisation

In practice either because of numerical approximation or experimental error, the element of the range on which the pseudo inverse is to act may not be known exactly. Thus for the answers obtained to make sense we need to ensure some continuity property with respect to these perturbations of the data. In order to do this, we here use the ideas of Tikhonov (Definition 2.35, Theorem 2.36) and introduce appropriate compactly embedded Hilbert spaces. Rather than attempt a formulation for the general problem we here present the two specific instances of most use to us, viz... .

a. recall (2.29), (2.30), (2.31). These can be regarded, with $M = \varepsilon^2 I_H$, as concerning a map $T:H \to \mathcal{Y}$ (given by $CS_t\, z_0$ in semi-group notation) and its regularisation $\tilde{T}_{\varepsilon,H}$. Suppose now we introduce a Hilbert space Z_1 compactly embedded in H. The operator $\tilde{T}_{\varepsilon,Z_1}$ will possess a bounded pseudo-inverse and have the continuity property (with respect to approximation in the output space $\mathcal{Y}$) which is described in Theorem 2.36. For future reference we make

<u>Definition 2.49</u>

The operator $\tilde{T}^{\dagger}_{\varepsilon,Z_1}$ described above, will henceforth be denoted by ${}_0\tilde{H}^{\dagger}_{\varepsilon,Z_1}$, as is consistent with the notation H_0.

b. Now we wish to consider a formulation involving model error. Define the map T by

$$\begin{aligned} &T : L^2(0,t_1;H) \times H \to L^2(0,t_1;H) \times \mathcal{Y} \\ &T ; (z,z_0) \to (z(\cdot) - S(\cdot)z_0 \,,\, Cz(\cdot)) \end{aligned} \qquad \ldots \qquad (2.37)$$

where we have committed some abuse of notation by mixing the semigroup and Lions formulations. The interpretation, however, is obvious. Note that C is now a map $C : L^2(0,t_1;H) \to \mathcal{Y}$; some circumstances may demand that $C : W(0,t_1) \to \mathcal{Y}$. We introduce, as before, a (Hilbert) space Z_1 compactly embedded in H. With an eye to the facts of Appendix 1 we make

Definition 2.50

By analogy with Definition 2.19 one takes

$$W_{Z_1}(0,t_1) = \{f \mid f \in L^2(0,t_1;Z_1), \frac{df}{dt} \in L^2(0,t,;Z_1^*)\}$$

with the norm

$$||f||_{W_{Z_1}(0,t_1)} = (\int_0^{t_1} ||f(s)||^2_{Z_1} ds + \int_0^{t_1} ||\frac{df}{dt}(s)||^2_{Z_1^*} ds)^{\frac{1}{2}} \quad .$$

Then by Appendix 1, $W_{Z_1}(0,t_1)$ is compactly embedded in $L^2(0,t_1;H)$. Thus $W_{Z_1}(0,t_1) \times Z_1$ is compactly embedded in $L^2(0,t_1;H) \times H$. Hence one can define $\tilde{T}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1}$ and obtain, as an immediate consequence of Theorem 2.36 the following...

Theorem 2.51

Consider the equation

$$\begin{aligned} z(t) - S(t)z_0 &= f_\varepsilon(t) \\ Cz(t) &= y_\varepsilon(t) \end{aligned} \qquad \ldots \qquad (2.38)$$

where $(f_\varepsilon,y_\varepsilon) \to (f,y)$ in $L^2(0,t_1;H) \times \mathcal{Y}$ as $\varepsilon \to 0^+$. Suppose that (f,y) lies in the range of $T|_{W_{Z_1}(0,t_1)\times Z_1}$; i.e. there exists (z^*,z_0^*) in $W_{Z_1}(0,t_1)\times Z_1$ such that

$$z^*(t) - S(t)z_0^* = f(t)$$

$$Cz^*(t) = y(t) \; .$$

Now let $(z^{\varepsilon}, z_0^{\varepsilon}) = \tilde{T}^{\dagger}_{\varepsilon, W_{Z_1}(0,t_1)\times Z_1}(0,0,f_{\varepsilon},y_{\varepsilon})$.

Then as $\varepsilon \to 0$, $(z^{\varepsilon}, z_0^{\varepsilon}) \to (z^*, z_0^*)$ with convergence in the norm of $L^2(0,t_1;H) \times H$.

Pf. by Theorem 2.36.

<u>Commentary</u>

It is worth making a number of remarks on the above results.

1. Finding a space Z_1 compactly embedded in H is often not difficult (see Appendix 2); it is also desirable, however, to make some natural choice. One can often take, for example, $Z_1 = V$ and thus $W_{Z_1}(0,t_1) = W(0,t_1)$ as previously defined.

2. In the case that $Z_1 = V$ we should expect to find some connection between the linear least squares problem and other notions of solution. Indeed, if $(f,y) \in$ range $(T|_{W(0,t_1)\times V})$ then $f \in W(0,t_1)$; and if one takes $f(0) = 0$ then $z^*(0) = z_0^*$; additionally if f has a representation as $f(t) = \int_0^t S(t-s)g(s)ds$ for some $g(\cdot) \in L^2(0,t_1;H)$ then we are dealing with a mild solution. Note, however, that the initial state z_0^* lies in V and not, as is usually the case, in H .

3. Often this last restriction (i.e. $z_0^* \in V$) is not too important. For not only is V dense in H , but also it is often physically desirable to recover an initial state in V rather than H . Remember that in general $z(\cdot) \in L^2(0,t_1;V)$ and so it is not possible to speak of its values in V other than in an almost everywhere sense.

4. Such a restriction, to look at a smoother type of solution, (i.e. $z_0 \in V$ rather than $z_0 \in H$) is the characteristic feature of Tikhonov regularisation techniques. Also typical is the fact that convergence is not obtained in the smoother space, but in the original rougher space (i.e. H rather than Z_1 ; $L^2(0,t_1;H) \times H$ rather than $W_{Z_1}(0,t_1) \times Z_1$) .

5. This treatment may appear to be excessively complicated. The following motivatory remarks are intended to justify its introduction. Many of the linear parabolic equations one would wish to handle throw up problems which are ill-posed in the sense of Definition 2.32. Moreover these problems are usually not "stable" with regard to approximation of data; comments on this notion of stability and a straightforward approach for use when it obtains will be found in Céa [1]. This desirable property is ensured for our purposes by using the approach of Tikhonov. The fact that the regularised pseudo-inverse maps into a compactly embedded space is an advantage in the fixed point formulations to come. Moreover we ensure that any algorithm based on using a sequence of linear approximate problems to arrive at the fixed point makes sense; as each linear problem is well posed and "stable" with respect to data perturbations.

Other observations

In the preceding, attention has been restricted to output operators $C : L^2(0,t_1;H) \to \mathcal{Y}$. This was done so as to obtain a natural setting for the Tikhonov regularisation. As was previously noted, it is more appropriate to consider $C : W(0,t_1) \to Y$; or more simply $C : L^2(0,t_1;V) \to Y$. These last two alternatives permit study of certain boundary or pointwise observations; though a useful, general theory of such systems remains to be constructed (for an indication of the difficulties, see Curtain Pritchard [1], Chapter 8). In view of the rudimentary nature of even the linear theory, these matters will, for the most part, be ignored in the present work. In the rest of this section we content ourselves with some remarks on these topics.

Recall the system studied in (2.29), (2.30)

$$\frac{dz}{dt} + A(t)z = f \qquad \text{in } \Omega \times]0,t_1[\qquad \ldots \qquad (2.39)$$

$$z = 0 \qquad \text{on } \Sigma = \partial\Omega \qquad \ldots \qquad (2.40)$$

$$z\big|_{t=0} = u \qquad \text{in } \Omega \qquad \ldots \qquad (2.41)$$

For a point $b \in \Omega$ one wishes to define a cost functional

$$J(u) = \int_0^{t_1} (z(b,t) - y_d(t))^2 dt + \varepsilon^2 ||u||_H^2 \quad \ldots \qquad (2.42)$$

where $y_d \in L^2(0,t_1;\mathbb{R})$, $\varepsilon > 0$. The sensible definition of this cost function, however, is not always possible.

Example 2.52

Take (2.39), (2.40), (2.41) and $\Omega \subset \mathbb{R}^1$, $A(t) = \Delta$ the Laplacian (i.e. $A(t)z = \frac{\partial^2 z}{\partial x^2}$) . Since $H_0^1(\Omega) \subset C^0(\Omega)$, $z(\cdot) \in L^2(0,t_1;H_0^1(\Omega))$ implies that $z(b,\cdot) \in L^2(0,t_1;\mathbb{R})$. Hence the cost functional (2.42) makes sense.

Example 2.53

Take (2.39), (2.40), (2.41) and Ω to be the unit ball in $\mathbb{R}^3$, $A(t) = \Delta$. Then it can be shown (see Lions [4]) that for $f \equiv 0$ there exists $u \in L^2(\Omega)$ with support on $\{x : ||x||_{\mathbb{R}^3} \leq 1\}$ so that $z(0,\cdot) \notin L^2(0,t_1;\mathbb{R})$.

One can make it so by restricting attention to a smaller class of admissible initial states. ∎

Thus in Example 2.52 our "initial state to output" framework (i.e. that which gives rise to ${}_0\tilde{H}^{\dagger}_{\varepsilon,Z_1}$ of Definition 2.49) is still valid. The framework including model error (i.e. (2.37)) is not valid as C is unbounded on $L^2(0,t_1;H)$. In Example 2.53 neither framework is valid; the "initial state to output" version can be recovered by restricting the set of admissible initial states to the (Hilbert) space

$$\mathcal{U} = \{u | u \in L^2(\Omega), z(b,\cdot;u) \in L^2(0,t_1;\mathbb{R})\}$$

where $\mathcal{U}$ has the norm

$$||u||_{\mathcal{U}} = (|u|^2_{L^2(\Omega)} + \int_0^{t_1} z(b,\cdot;u)^2 dt)^{\frac{1}{2}}$$

With this restriction, however, it is not clear whether one can find natural compactly embedded subspaces of $\mathcal{U}$; thus the Tikhonov formulation cannot be used. One can, however, construct a well-defined problem (2.39), (2.40), (2.41), (2.42) whose solution is given by ${}_0\tilde{H}^{\dagger}_{\varepsilon,\mathcal{U}}(0,y_d)$ where ${}_0\tilde{H}_{\varepsilon,\mathcal{U}}$ denotes the map

$$ {}_0\tilde{H}_{\varepsilon,\mathcal{U}} : \mathcal{U} \to \mathcal{U} \times \mathcal{Y} $$

$$ {}_0\tilde{H}_{\varepsilon,\mathcal{U}} : u \to (u, z(b,\cdot;u)) $$

and $\mathcal{U}$ on the right-hand side is, as is usual, given the equivalent norm $\varepsilon^2||\cdot||_{\mathcal{U}}$, $\varepsilon > 0$. Thus we can, at least theoretically, pose this problem in a linear least squares context; as the use of "theoretically" implies there are a number of numerical and computational questions whose answers are unknown (e.g. how to characterise $\mathcal{U}$, how sensitive is the result to data errors).

It is obvious that the study of boundary observation (and control) offers many possible research problems. One might, for instance, investigate the approximation of $C \in \mathcal{L}(W(0,t_1);\mathcal{Y})$ (or $\mathcal{L}(L^2(0,t_1;V);\mathcal{Y})$) by $C_\varepsilon \in \mathcal{L}(L^2(0,t_1;H);\mathcal{Y})$ and appropriate (if any) notions of convergence. The main object of this thesis is to indicate some ways of looking at non linear problems and not to find the best possible setting for linear systems with pointwise observations (or control). The above remarks should be taken as caveats in respect of the thesis' generality.

CHAPTER III : A Non-Linear Theory

Summary

This chapter begins with an introduction to non linear partial differential equations. This section is necessarily brief but aims to indicate some justification for the approach to reconstruction (and, by analogy control) problems adopted later in the chapter. Further details will be found in Haraux [1], Henry [1], Lions [3]. In these later sections we indicate the use of the linear part, in conjunction with some fixed point theorems, to construct and prove theorems about a class of non-linear systems. Algorithmic, and other, aspects of these results are discussed in later chapters.

3.1 Non linear evolution equations

The problems encountered in the study of non linear *ordinary* differential equations are numerous. Indeed, this is a currently active area of research. Some pathological (at least, by comparison with the linear case) phenomena which occur are indicated by standard examples, viz... .

Example 3.1

The solutions of linear evolution equations, such as are studied in Chapter II, can usually be extended for all positive time; this is certainly true of the semigroup formulation. The same does not hold for non-linear ordinary differential equations: consider

$$\frac{dz}{dt} = z^2 \quad ; \quad z(0) = a > 0 \qquad \ldots \qquad (3.1)$$

then the only solution is

$$z(t) = \frac{a}{1-at} \quad \text{for} \quad 0 \le t < \frac{1}{a} \qquad \ldots \qquad (3.2)$$

Obviously this does not exist for all $t > 0$, but only for t sufficiently small. This problem of rapid growth, or "blow-up", is fundamental and cannot be excised by additional smoothness or other such assumptions.

Example 3.2

The solutions of well posed (Hadamard) linear evolution equations are unique. Again this is not so even for non-linear ordinary differential equations: consider (for $\alpha \in]0,1[$)

$$\frac{dz}{dt} = |z|^\alpha \quad ; \quad z(0) = 0 \qquad \ldots \qquad (3.3)$$

this has the obvious solution $z(t) = 0 \;\; \forall t \ge 0$; and also infinitely many other solutions; for any $\tau > 0$ take

$$x(t) = \begin{cases} 0 & \text{for} \quad 0 \le t \le \tau \\ p^{-p}(t-\tau)^p & \text{for} \quad t \ge \tau \end{cases}$$

where $p = \frac{1}{1-\alpha}$.

In this case, the problem is one of insufficient smoothness as the right hand side of the equation is not Lipschitz continuous in z .

Additionally, of course, it is possible to write down systems of non-linear ordinary differential equations which do not have any solutions. Thus one is faced by non-existent, non-global, non-unique solutions; though not all at once.

Infinite dimensional case : formulation

Given a non linear partial differential equation it is, at the very least, desirable to have an appropriate notion of solution, with which questions of existence and uniqueness can be studied. As in the linear case one has the usual problems of interpreting a formal expression: that is, one can search for strong, mild or weak solutions. Moreover, non linearities often map one outside a given domain space. This is not an unusual occurrence in practical problems.

Example 3.3

Consider the nonlinear map $N : z \to z^2$ where z is a real valued function on the interval $[0,1]$. Suppose that the desired range space is $L^2[0,1]$. Then an appropriate domain space for N would be $L^4[0,1] \subset L^2[0,1]$.

Another problem is that global Lipschitz conditions often do not apply.

Example 3.4

Consider $N:L^4[0,1] \to L^2[0,1]$, $Nz = z^2$ as above, then there does not exist $k > 0$ such that

$$||z_1^2 - z_2^2||_{L^2} \leq k||z_1 - z_2||_{L^4} \quad \forall z_1, z_2 \in L^4 .$$

To provide some background we recall the classical result.

Theorem 3.5

Let Z be a Banach space, $[0,t_1]$ the time interval of interest and $f : [0,t_1] \times Z \to Z$ be continuous. Assume also that f is locally Lipschitzian in z, uniformly with respect to t. Then $\forall t_0 \in [0,t_1[$ and each $z_0 \in Z$ $\exists$ $\delta > 0$ and a unique strong solution on $[t_0,t_0+\delta[$ of the Cauchy problem

$$\frac{dz}{dt} = f(t,z(t)) \ , \quad z(t_0) = z_0$$

(recall that by strong solution we mean a $z(\cdot)$ explicitly satisfying the differential equation (and initial condition)).

Pf. Define

$$F : z \to z_0 + \int_{t_0}^{t} f(s,z(s))ds$$

$$V_\delta = C([t_0,t_0+\delta[;X)$$

$$B_\alpha = \{z \in V_\delta : ||z - z_0||_Z \leq \alpha\}$$

for α small enough

$$\sup_{\substack{s\in[0,t_1] \\ ||z-z_0||_Z\leq\alpha}} ||f(s,z)||_Z \leq M < \infty .$$

If $M.\delta \leq \alpha$ then $F : B_\alpha \to B_\alpha$. Also, by the local Lipschitz assumption, if α,δ are small enough, F is a strict contraction from B_α to B_α . Hence the result by the contraction mapping theorem (see Appendix 4). ∎

It can also be shown that if f is Lipschitzian in X (other assumptions as Theorem 3.5) then one can take $t_0 = 0$ and $\delta = t_1$. As noted in Chapter I many systems can be viewed as a nonlinear perturbation of a linear part. Certainly this is true locally in time - by using some form of approximation - and often, as we have seen we may only be able to define solutions locally in time and (state) space in any case. This type of non-linear equation is also more tractable analytically - for one can trade off the properties of linear and non linear parts against each other; for instance a smoothing property of the semigroup generated by the linear part against the "roughness" of the nonlinear part. Consider the equation

$$\dot{z} = Az + f(t,z(t)) \quad \ldots \quad (3.4)$$

where A generates a semigroup $S(t)$ on a Banach space Z . Hereafter an equation such as (3.4) will be called semi-linear. Concerning a mild solution of (3.4), valid for all $z_0 \in Z$, there is the result of Segal [1].

<u>Theorem 3.6</u>

If f satisfies the hypotheses of Theorem 3.5 then $\forall z_0 \in Z$ $\exists\, \delta > 0$ such that the (nonlinear) integral equation

$$z(t) = S(t)z_0 + \int_0^t S(t-s)f(s,z(s))ds \quad \ldots \qquad (3.5)$$

has an unique solution in $C([0,\delta];Z)$.

Pf. As Theorem 3.5.

The assumptions on the nonlinearity can be restrictive; consider for instance

Example 3.7

Here we are concerned to formulate the nonlinear Schrödinger equation

$$i\frac{dz}{dt} + \Delta z = g(|z|^2)z \quad \ldots \qquad (3.6)$$

in the space $Z = H^2(\mathbb{R}^n;\mathbb{C})$ where $n = 2,3$. Now Z is an algebra included in $L^\infty(\mathbb{R}^n;\mathbb{C})$ (see also Appendix 1); this is a consequence of (D denotes $\frac{\partial}{\partial x}$)

$$|Dz|_{L^4} \leq C|z|^{\frac{1}{2}}_{L^\infty}|D^2z|^{\frac{1}{2}}_{L^2} \quad \ldots \qquad (3.7)$$

(which is true for $z \in H^2(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n)$) and the fact that H^2 embeds in L^∞ (for $n = 2,3$). In order that $f(Z) \subset Z$ it is enough to demand that

$$g \in C^0(\mathbb{R}^+), g', sg''(s) \in L^\infty_{loc}(\mathbb{R}^+)$$

This gives, for instance, that

$$i\frac{dz}{dt} + \Delta z = k|z|^2z \quad \ldots \qquad (3.8)$$

has a local solution in $H^2(\mathbb{R}^n)$ $(n = 2,3)$.

Thus far we have seen how an integral, or mild, formulation for non linear partial differential equations can be constructed. Local existence and uniqueness results are then proven using fixed point arguments. Now it is appropriate to investigate the effects of varying the assumptions on the non-linearity, f, and the semigroup $S(\cdot)$ in (3.5) with a view to proving more easily applicable theorems.

Further results

Often $f(t,z)$ in (3.4) is differentiable with respect to t and z; this property can be used to relax other assumptions on $f(\cdot,\cdot)$.

Theorem 3.8

Let f be differentiable from $[0,t_1] \times D(A)$ into Z. Assume also that the function

$$g(t;u,v) = \frac{\partial f}{\partial t}(u) + \frac{\partial f}{\partial u}(u,v) \quad \ldots \quad (3.9)$$

is locally Lipschitz : $D(A) \times Z \to Z$ (uniformly in t). Then for $z_0 \in D(A)$ the equation (3.4) has an unique strong solution $z(\cdot) \in C([0,\delta];D(A))$ for some $\delta > 0$.

Pf. see Segal [1]. The proof proceeds by setting $v = u_t$ and studying the system

$$\begin{aligned} u_t &= v \\ v_t &= Av + g(t;u,v) \end{aligned} \quad \ldots \quad (3.10)$$

treated in the space $D(A) \times Z$ as a perturbation of the linear system

$$\begin{aligned} u_t &= v \\ v_t &= Av \end{aligned} \quad \ldots \quad (3.11)$$

Obviously (3.11) generates a (linear) semigroup

$$L(t)\ (u_0,v_0) = (u_0 + \int_0^t S(s)v_0 ds,\ S(t)v_0) \quad \ldots \quad (3.12)$$

Versions of the preceding theorems are used to complete the result. ∎

A recent example from the literature showing the use of this theorem is

Example 3.9

Arising from a mathematical model of liquid crystal behaviour studied by Dias we have the system ...

$\Omega \subset \mathbb{R}^3$; Ω bounded with a regular boundary Γ

$Z = (L^2(\Omega))^3$, $h(\cdot) \in C^1(\mathbb{R}^+;\mathbb{R}^3)$

$$\frac{\partial z}{\partial t} - \Delta z + (z.h(t))^2 z - (z.h(t))h(t) = 0$$

$$z.n_1 = 0 \text{ on } \Gamma \quad \ldots \quad (3.13)$$

$$\frac{\partial z}{\partial n_1} \times n_1 = 0 \quad \text{on } \Gamma$$

$$D(A) = \{z \in (H^2(\Omega))^3 : z.n_1 = 0 \text{ on } \Gamma,\ \frac{\partial z}{\partial n_1} \times n_1 = 0 \text{ on } \Gamma\}$$

where n_1 is the normal to Γ .

Then Theorem 3.8 may be applied to give a local existence and uniqueness result for any $z_0 \in D(A)$. See Dias (additional references).

As has already been stated, if the semigroup involved has a "smoothing" action then it may be used to "smooth" the non-linearity; these comments are most appropriate in regard to linear parabolic parts and the corresponding analytic semigroups - with reference to Appendix 2 (for definition of sectorial, Z^α etc.) we may state

Theorem 3.10

Let $-A$ be a sectorial operator, $\alpha \in [0,1[$ and $f : U \to Z$ where U is an open subset of $\mathbb{R}^+ \times Z^\alpha$. Assume also that $f(t,z)$ is locally Lipschitz (in (t,z)) ; then for any $(t_0,z_0) \in U, \exists\ \delta = \delta(t_0,z_0) > 0$ such that (3.5) has an unique solution $z(\cdot) \in C([t_0,t_0 + \delta];Z^\alpha)$.

Pf. exactly analogous to that of Theorem 3.5

In this case, by assuming additional smoothness on f (most usually that $t \to f(t,z)$ is locally Hölder continuous) one can show the existence and uniqueness of strong solutions; i.e. solutions which satisfy the differential equation (3.4). In the present work, as has already been noted in Chapter II, we prefer to work with mild solutions which involve spaces more appropriate to our formulation of the estimation (and control) problems. Some examples from the literature will illustrate the use of the above Theorem.

Example 3.11

Consider (cf. Navier-Stokes) the system

$$\frac{\partial z}{\partial t} + z\,\frac{\partial z}{\partial x} = \frac{\partial^2 z}{\partial x^2} + f(t,z)$$
$$z(0,t) = z(\pi,t) = 0 \qquad \ldots \quad (3.14)$$

where $f: \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ is locally Hölder continuous in t and locally Lipschitz in z with $|f(t,v)| \leq g(t,|v|)$ where g is continuous and increasing in the second variable.

Now take $Z = L^2(0,\pi)$; as in Example 2.15 we have $A \equiv -\frac{d^2}{dx^2}$ with domain $H^2(0,\pi) \cap H_0^1(0,\pi)$ and $D(A^{\frac{1}{2}}) = H_0^1(0,\pi)$. It can now be shown that

$$F : \mathbb{R}^+ \times H_0^1(0,\pi) \to L^2(0,\pi)$$

where $F(t,\phi)(x) = -\phi(x)\phi'(x) + f(t,\phi(x))$ $0 < x < \pi$ satisfies the hypotheses of Theorem 3.10. Thus (3.14) has a mild solution in $C([0,\delta];H_0^1(0,\pi))$ for some $\delta > 0$.

Example 3.12

Consider the system

$$\frac{\partial z}{\partial t} = \frac{\partial^2 z}{\partial x^2} + z - z^2 \qquad \ldots \qquad (3.15)$$
$$z(0,t) = z(1,t) = 0$$

Suppose $z(x,0) = z_0(x) \geq 0$ on $0 \leq x \leq 1$ where $z_0 \in H_0^1(0,1)$. Then a standard maximum principle argument (see, for example, Protter and Weinberger [1]) gives that $z(\cdot,t) \geq 0$ for all times t , in its interval of existence; moreover, the solution can be proven to exist for all $t \geq 0$.

One can go on to study many other aspects of the solutions in Theorem 3.10; for instance smoothness of parameter dependence, asymptotic stability of certain solutions (e.g. zero solution in (3.15)), periodic solutions. These directions will not be pursued here; although they are obviously of relevance to certain control and estimation applications.

So far, we have considered mild solutions lying in $C([0,t_1];Z)$ for some t_1 and Z. The integral formulation (3.5), however, can be used to define notions of solution lying in more general spaces (e.g. $L^r([0,t_1];Z)$ for some r,t_1 and Z). Thus one can hope to include larger classes of non-linearities; and also provide a framework more suited to certain classes of control, estimation and optimisation problems where such spaces occur naturally. Much work has been done in this direction; the next theorem gives an example due to Ichikawa-Pritchard [1]. See also Kato-Fujita, Weissler (additional references).

Theorem 3.13

Let V,Z_1,Z_2 be Banach spaces with $V \subset Z_1$; and $a,b,p_1,p_2,q,r,s,t_1 \in \mathbb{R}^+$ satisfying $p_1 \geq r \geq 1$, $p_2 \geq q \geq 1$ $s \geq 1$, $\frac{1}{r} = \frac{1}{q} + \frac{1}{s} - 1$. Assume also

i) $S(t) \in L(Z_1,V) \cap L(Z_2,V)$ for $t > 0$ with

$$||S(t)z||_V \leq g_1(t)||z||_{Z_1} \quad t > 0 \quad \forall z \in Z_1 \text{ and}$$

$$||S(t)z||_V \leq g_2(t)||z||_{Z_2} \quad t > 0 \quad \forall z \in Z_2$$

where $g_1 \in L^{p_1}(0,t_1;\mathbb{R}), g_2 \in L^{p_2}(0,t_1;\mathbb{R})$

ii) $f : V \to Z_2$ is such that $z(\cdot) \in B_a$ defined by

$B_a \equiv \{z(\cdot) \in L^r(0,t_1;V) : ||z(\cdot)||_{L^r(0,t_1;V)} \leq a\}$ implies that

$f(z(\cdot)) \in L^s(0,t_1;Z_2)$ and $\exists\, b : ||f(z(\cdot))||_{L^s(0,t_1;Z_2)} \leq b$

iii) $||f(z(\cdot))-f(\hat{z}(\cdot))||_{L^s(0,t_1;Z_2)} \leq k(||z(\cdot)||_{L^r(0,t_1;V)},$

$||\hat{z}(\cdot)||_{L^r(0,t_1;V)})||z(\cdot)-\hat{z}(\cdot)||_{L^r(0,t_1;V)}$

where $k:\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$, continuous, symmetric and such that $k(\theta_1,\theta_2) \to 0$ as $(\theta_1,\theta_2) \to (0,0)$

iv) for $z(\cdot)$, $\hat{z}(\cdot) \in B_a$

$$||g_2||_{L^q(0,t_1;\mathbb{R})} \cdot k(||z(\cdot)||_{L^r(0,t_1;V)}, ||z(\cdot)||_{L^r(0,t_1;V)}) < 1$$

v) for $z_0 \in Z_1$

$$||g_1||_{L^r(0,t_1;\mathbb{R})} ||z_0||_{Z_1} + ||g_2||_{L^q(0,t_1;\mathbb{R})} b \leq a$$

Then there exists a solution of (3.5) in $L^r(0,t_1;V)$ (unique in B_a).

Pf. The objective is to apply the contraction mapping theorem (see Appendix 4) to the map $\Phi:(\Phi z)(t) = S(t)z_0 + \int_0^t S(t-\tau)f(z(\tau))d\tau$. First we show that $\Phi:B_a \to B_a$.

By i) and ii)

$$||(\Phi z)(t)||_V \leq g_1(t)||z_0||_{Z_1} + \int_0^t g_2(t-\tau)||f(z(\tau))||_{Z_2} \, d\tau .$$

Viewing the second term on the right hand side as a convolution ($g_2 \in L^q(0,t_1;\mathbb{R})$, $||f(z)||_{Z_2} \in L^s(0,t_1;\mathbb{R})$) ; see Dieudonné (additional references) p.291. Hence we obtain

$$||\Phi z||_{L^r(0,t_1;V)} \le ||g_1||_{L^r(0,t_1;\mathbb{R})} ||z_0||_{Z_1} +$$

$$+ ||g_2||_{L^q(0,t_1;\mathbb{R})} ||f(z)||_{L^s(0,t_1;Z_2)} \quad .$$

From ii) and v) we conclude

$$||\Phi z||_{L^r(0,t_1;V)} \le a \ . \quad \text{Thus} \quad \Phi : B_a \to B_a \ .$$

Next, it is required to show that Φ is a contraction on B_a . Consider $\Phi z - \Phi\hat{z}$ for $z,\hat{z} \in B_a$; by the same "convolution technique" as used above one obtains

$$||\Phi z - \Phi\hat{z}||_{L^r(0,t_1;V)} \le ||g_2||_{L^q(0,t_1;\mathbb{R})} ||f(z)-f(\hat{z})||_{L^s(0,t_1;Z_2)} \quad .$$

By using iii), iv)

$$||\Phi z - \Phi\hat{z}||_{L^r(0,t_1;V)} \le K \, ||z-\hat{z}||_{L^r(0,t_1;V)}$$

where $K : 0 < K < 1$.

Thus, by the contraction mapping theorem, Φ has an unique fixed point in B_a . ∎

The mild solution used in this theorem (i.e. in $L^r(0,t_1;V)$ can be related to that used in the preceding theorems (i.e. $C(0,t_1;V)$) by the following.

Corollary 3.14

Suppose, in addition to the hypotheses of Theorem 3.13, that $S(t) \in L(Z_2, Z_1)$ for $t > 0$ and satisfies

$$||S(t)z||_{Z_1} \leq g_3(t)||z||_{Z_2} \qquad \forall z \in Z_2$$

where

$$g_3 \in L^{p_3}(0, t_1; \mathbb{R})$$

and

$$p_3 : \frac{1}{p_3} + \frac{1}{s} = 1 .$$

Then the solution proven to exist in $L^r(0,t_1;V)$ by Theorem 3.13, also lies in $C(0,t_1;V)$.

Pf. see Ichikawa-Pritchard [1].

The local Lipschitz condition iii) in Theorem 3.13 is more general than is usual since two different spaces are used. In the case that these spaces are distinct, the contraction property on $L^r(0,t_1;V)$ would be expected to result from some smoothing action of the semigroup $S(t)$. With the aid of the above formulation and further regularity results one can investigate solutions which are global in time. Typically one obtains a ball of initial statessuch that solutions starting there can be extended for all time. The size obtained for this ball makes precise the standard "for $||z_0||$ sufficiently small..." statements. All these aspects are studied in Ichikawa-Pritchard [1]. For future reference, this section concludes with an example drawn from this paper.

Example 3.15

Recall Example 2.13 with $\alpha = 0$ and a non-linearity $f_1(z,\dot{z})$, that is ...

$$\ddot{z} + Az + f_1(z,\dot{z}) = 0 \qquad \dots \qquad (3.16)$$
$$\dot{z}(0) = z_1 \; ; \; z(0) = z_0$$

Let A be as in Example 2.13, then by augmenting in the standard fashion we obtain a semigroup $S(t)$ on the product space $D(A^{\frac{1}{2}}) \times H$. The non-linearity is taken to be such that

$$F : D(A^{\frac{1}{2}}) \times H \to D(A^{\frac{1}{2}}) \times H$$
$$: (z,\dot{z}) \mapsto (0 \; , \; f_1(z,\dot{z})) \; .$$

The semigroup in this case does *not* smooth the space $D(A^{\frac{1}{2}}) \times H$ (as has been previously stated, this is a typical feature of hyperbolic, as opposed to parabolic problems). However, in this case we can take advantage of the special structure of F.

Suppose that $f_1 : D(A^{\frac{1}{2}}) \times H \to H$ satisfies

$$||f_1 w - f_1 \hat{w}||_H \le k(||w||_{D(A^{\frac{1}{2}})\times H}, ||\hat{w}||_{D(A^{\frac{1}{2}})\times H}) \cdot ||w-\hat{w}||_{D(A^{\frac{1}{2}})\times H}$$

then

$$||Fw-F\hat{w}||_{D(A^{\frac{1}{2}})\times H} \le k(||w||_{D(A^{\frac{1}{2}})\times H}, ||\hat{w}||_{D(A^{\frac{1}{2}})\times H}) \cdot ||w-\hat{w}||_{D(A^{\frac{1}{2}})\times H} \; .$$

Hence, with $V = Z_1 = Z_2 = D(A^{\frac{1}{2}}) \times H$, Theorem 3.13 applies, with $k(.,.) < 1$ on B_a, for any $r \geq 1$. By the corollary the solution obtained is also in $C(0,t_1;D(A^{\frac{1}{2}}) \times H)$.

Example 3.16

By analogy with 2.14, this is a special case of the preceding. Consider the following system for $x \in \Omega \subset \mathbb{R}^3$ where Ω is open bounded, subject to the appropriate smoothness conditions on $\partial\Omega$.

$$\begin{aligned} &z_{tt} = z_{xx} + f_1(z,z_t) \\ &z\big|_{\partial\Omega} = 0 \qquad \ldots \qquad (3.17) \\ &\dot{z}(0) = z_1 \;;\quad z(0) = z_0 \end{aligned}$$

We take $H = L^2(\Omega;\mathbb{R})$ (cf. Example 2.14), A defined as in Example 2.14,

$$D(A) = H^2(\Omega;\mathbb{R}) \cap H_0^1(\Omega;\mathbb{R})$$

and so $D(A^{\frac{1}{2}}) = H_0^1(\Omega)$. Candidate non-linearities

$f_1 : D(A^{\frac{1}{2}}) \times H \to H$ are

$f_1 : (z,z_t) \to z^\alpha$ for $1 \leq \alpha < 3$ (see Appendix 1)

or $f_1 : (z,z_t) \to D^\alpha z||z||^p_{H^1}$ $|\alpha| \leq 1$, $\forall p \geq 1$

where D^α is a differential operator.

Again from the results in Appendix 1, if $\Omega \subset \mathbb{R}$ then one can consider

$$f_1 : (z,z_t) \to z^\alpha \qquad \text{for } 1 \leq \alpha < \infty$$

or $$f_1 : (z,z_t) \to z^\alpha z_t \qquad \text{for } 1 \leq \alpha < \infty \;;$$

and hence impose conditions so as to ensure satisfaction of the local Lipschitz and contraction requirements.

Note that the emphasis in these last results (Theorem 3.13, Corollary 3.14), is on solutions which are "local in initial states" as well as local in time. That is, one searches for solutions which may only be defined for initial states in some ball, whose size may be proscribed. This differs markedly from previous work where one is concerned to define solutions globally. The need for such restrictions will now be indicated.

Pathology of solutions

In (3.1), (3.2) we have displayed an example of "blow-up" for a non-linear ordinary differential equation. The same phenomenon also occurs for non-linear partial differential equations.

Example 3.17

$$\frac{\partial z}{\partial t} = \frac{\partial^2 z}{\partial x^2} + z^3 \qquad 0 < x < \pi, \ t > 0$$

$$z(0,t) = z(\pi,t) = 0$$

$$z(x,0) = z_0(x) \ .$$

If $||z_0||_{H_0^1(0,\pi;\mathbb{R})}$ is sufficiently small it can be shown that the (strong) solution $z(\cdot,t)$ exists $\forall t > 0$ (and even tends to zero as $t \to \infty$) ; see, for instance, Henry [1]. Suppose now that $||z_0||_{H_0^1(0,\pi;\mathbb{R})}$

is not small; in particular, that

$$z_0(\cdot) \geq 0 \quad \text{on} \quad [0,\pi]$$

and

$$\int_0^\pi z_0(x) \sin x \, dx > 2 \quad .$$

Arguments based on the maximum principle (see Protter-Weinberger [1]) show that $z(x,t) \geq 0$ for $x \in [0,\pi]$, t on the interval of existence. Now set

$$g(t) = \int_0^\pi z(x,t) \sin x \, dx$$

so

$$\frac{dg}{dt} = - g + \int_0^\pi z^3(x,t) \sin x \, dx \quad .$$

Hölders inequality gives (using $(\sin x)^{1/3}.(\sin x)^{2/3}$)

$$g(t) \leq 2^{2/3} \left(\int_0^\pi z^3(x,t) \sin x \, dx\right)^{1/3}$$

and thus the differential inequality

$$\frac{dg}{dt} \geq - g + \frac{1}{4} g^3 \qquad \text{for} \quad t > 0$$

$$g(0) = \int_0^\pi z_0(x) \sin x \, dx > 2 \quad .$$

This differential inequality can be used to show that $g(t) \to + \infty$ in finite time (in fact: at, or before, $\frac{1}{2}\log((g(0)+2)/(g(0)-2)))$. Further analysis of this example can be found in Ichikawa-Pritchard [1] which gives an estimate for the region of asymptotic stability.

The use of techniques such as those above indicates either a. that the solution blows up, or b. that the solution has a maximal interval of existence strictly less than the blow-up time. From our present, pragmatic, point of view both these phenomena will be regarded as "solution pathologies". The following two theorems indicate for two archetypal equations (the wave and heat equations) when such pathologies occur. See Ball, [1], for a detailed discussion of these results. Both these theorems are stated with the understanding that reasonable assumptions have been to ensure the existence of a solution; perhaps only locally in time and for small initial data.

Theorem 3.18 (John)

Consider the non-linear wave equation in $\mathbb{R}^3$

$$z_{tt} - \Delta z = f(t,z)$$
$$t \geq 0 \; , \; x \in \mathbb{R}^3 \quad \ldots \quad (3.18)$$

subject to $f(t,s) \geq b|s|^p$ where $b > 0$ and with compactly supported initial data. If $1 < p < 1+\sqrt{2}$ any solution of (3.18) is "pathological" as defined above. This condition is sharp in that if $f(t,x,s) = |s|^p$ with $p > 1 + \sqrt{2}$ then the solution exists for all time, as long as the initial data is sufficiently small.

Pf. see John [1].

This result has been extended by a number of authors (e.g. Glassey, Kato). For instance Sideris has shown that, on $\mathbb{R}^1$, pathologies develop $\forall p > 1$ (here the proof turns on the fact that the solution of $z_{tt} = z_{xx}$ with same initial data (as the non-linear problem) does not decay uniformly to zero).

Theorem 3.19 (Fujita)

Consider the non-linear heat equation

$$\begin{aligned} & z_t - \Delta z = f(t,z) \\ & t \geq 0 \ , \ x \in \mathbb{R}^n \qquad \ldots \qquad (3.19) \\ & z(x,0) = z_0(x) \geq 0 \end{aligned}$$

subject to $f(t,s) \geq b|s|^p$ where $b > 0$ and with compactly supported initial data.

If $1 < p \leq \frac{n+2}{n}$ any solution of (3.18) is "pathological", as defined above.

This condition is sharp in that if $f(t,x,s) = s^p$ with $p > \frac{n+2}{n}$, then the solution exists for all time, as long as the initial data is sufficiently small.

Pf. see Fujita [1].

These results point to some fundamental restrictions on modelling with non-linear partial differential equations; and that, in general, it is only reasonable to ask about solutions defined for sufficiently small initial data and intervals of time.

Philosophical issues

Since we hope ultimately, even if not in this thesis, to address control and estimation problems of practical significance, it is worth asking "given a system of non-linear partial differential equations derived by modelling physical reality, what is implied by the restrictions on the (mathematical) solution noted above?" For instance we might not expect a well-behaved physical reality to blow-up at, or fail to exist after, some time (this is not a Berkeleyan justification for the C.N.D.). Additionally, if we have a good model we would surely expect it to be valid for a variety of initial states.

The justification adopted by the author is that all modelling involves approximation. That is, to arrive at any model, assumptions descriptive of some particular regime of operation have been made. The model cannot reasonably be expected to provide a good approximation outside of this regime. "Blow-up" may indicate a transition to some qualitatively different behaviour. Such transitions are not unknown in non-linear systems; probably the most well-known example is the onset of turbulence in fluid flow. From this viewpoint it is not unreasonable to produce models which have only "local meaning" (i.e. local in time and states). The above restrictions on the notion of solution are thus consistent with this interpretation of modelling.

3.2 Reconstruction for non-linear systems

This section considers the problem of reconstructing the state of a system, governed by a non-linear evolution equation, given the available measurements. The approach is entirely analogous to that of the preceding section. There the properties of the linear part, the semigroup, were used to construct a representation - the mild solution, by variation of constants - and to assist in the investigation of its properties (viz. the trade off between the smoothing action of the semigroup and the non-linearity). In this section our knowledge (see Chapter II) of linear state reconstruction is used to cast the non-linear problem as a fixed point one. Standard fixed point theorems are then used to obtain the desired results. First we provide some justification for the formulation.

Formulation

Here, and subsequently, the non-linearity is taken to be autonomous, i.e. time independent; it will thus be a function of the state alone. This assumption is not a major restriction - it can easily be removed - it merely serves to simplify the presentation. More contentious, perhaps, is the assumption that the linear part also is time invariant. This, again, is not intrinsic, for the methods can straightforwardly be extended to cover the evolution operator case. It is rare, however, for linearisation to yield a time invariant linear system; in general one linearises about some time-varying trajectory. The present aim is to demonstrate the techniques rather than prove the most general theorem possible.

Suppose, then, that we consider the evolution equation

$$\dot{z} = Az + f(z) \quad \ldots \quad (3.20)$$

with observations

$$y = Cz \quad \ldots \quad (3.21)$$

where C is a linear output operator, and A generates a semigroup $S(t)$. The linear part

$$\begin{aligned} \dot{z} &= Az \\ y &= Cz \end{aligned} \quad \ldots \quad (3.22)$$

gives rise to the "initial state to output" operator on $[0,t_1]$

$$H_0 : z_0 \to CS(\cdot)z_0 \quad \ldots \quad (3.23)$$

Suppose that the linear system (3.22) is continuously initially observable with respect to some space $\mathcal{Y}$; then H_0^{-1} (or H_1 , as it is called in the discussion preceding Definition 2.40) exists. Consider now a mild solution of (3.20), given, if it exists, by

$$z(t) = S(t)z_0 + \int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (3.24)$$

then operating on both sides by C one has

$$y(t) = CS(t)z_0 + C\int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (3.25)$$

Thus by rearranging (3.25) we have

$$z_0 = H_0^{-1}(y(\cdot) - C\int_0^{\cdot} S(\cdot - s)f(z(s))ds) \quad \ldots \quad (3.26)$$

and substitution of (3.26), for z_0 in (3.24), gives

$$z(t) = S(t)H_0^{-1}(y(\cdot) - C\int_0^{\cdot} S(\cdot - s)f(z(s))ds) + \int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (3.27)$$

The right-hand side of (3.27) is used to define a map Φ acting on the "trajectory space". Note that $C(\Phi(z))(t) = y(t)$. Hence a fixed point of Φ will be consistent with the original dynamics and output equation. The rest of this chapter will be concerned with making this general approach rigorous, i.e. proving existence (and, in some cases, uniqueness) for such fixed points. The maps, whose fixed points are sought, are constructed from some known linear reconstruction problem, whilst regarding the non-linearity as a known perturbation. When the real non-linearity is inserted one obtains a map whose fixed points will be trajectories consistent with the original non-linear equations. This might be regarded as the estimation (or, by duality, control) version of the so-called "Schauder linearisation procedure". To prove results in this area we need, as in section 3.1, conditions on the linear part, the non-linearity and their interaction.

Contraction mapping result

Take the map Φ defined by

$$(\Phi(z))(t) = S(t)H_0^{-1}(y(\cdot) - C\int_0^{\cdot} S(\cdot - s)f(z(s))ds) + \int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (3.28)$$

where H_0 is taken as mapping from Z_1 to $\mathcal{Y}$, and is invertible i.e. linear part as a system on Z_1 with output in $\mathcal{Y}$ is continuously initially observable.

The following result is patterned after Theorem 3.13.

Theorem 3.20

Let V, Z_1, Z_2 be Banach spaces with $V \subset Z_1$; and $a, K, p_1, p_2, q, r, R, s, t_1 \in \mathbb{R}^+$ satisfying $p_1 \geq r \geq 1$, $p_2 \geq q \geq 1$, $s \geq 1$, $\frac{1}{r} = \frac{1}{q} + \frac{1}{s} - 1$. Assume also

i) $S(t) \in L(Z_1,V) \cap L(Z_2,V) \qquad t > 0$

$$||S(t)z||_V \leq g_1(t)||z||_{Z_1} \qquad t > 0 \quad \forall z \in Z_1 \quad \text{and}$$

$$||S(t)z||_V \leq g_2(t)||z||_{Z_2} \qquad t > 0 \quad \forall z \in Z_2$$

where $g_1 \in L^{p_1}(0,t_1;\mathbb{R})$, $g_2 \in L^{p_2}(0,t_1;\mathbb{R})$.

ii) $R > 0$ is such that

$$||C\int_0^{\cdot} S(\cdot-\tau)z(\tau)d\tau||_{\mathcal{Y}} \leq R||z||_{L^s(0,t_1;Z_2)}$$

where $\mathcal{Y}$ is the space for which continuous initial observability holds.

iii) $f : V \to Z_2$ is such that

$$||f(z(\cdot))-f(\hat{z}(\cdot))||_{L^s(0,t_1;Z_2)} \leq k(||z(\cdot)||_{L^r(0,t_1;V)},$$

$$||\hat{z}(\cdot)||_{L^r(0,t_1;V)})||z(\cdot)-\hat{z}(\cdot)||_{L^r(0,t_1;V)}$$

where $k : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, symmetric and such that $k(\theta_1,\theta_2) \to 0$ as $(\theta_1,\theta_2) \to (0,0)$

iv) taking $B_a = \{z \in L^r(0,t_1;V) : ||z||_{L^r(0,t_1;V)} \leq a\}$

such that for $z,\hat{z} \in B_a$

$$(R||H_0^{-1}||_{L(Y,Z_1)}||g_1||_{L^r(0,t_1;\mathbb{R})} + ||g_2||_{L^q(0,t_1;\mathbb{R})}) \cdot$$

$$k(||z||_{L^r(0,t_1;V)}, ||\hat{z}||_{L^r(0,t_1;V)}) \leq K < 1$$

Then: the state of the system described by (3.24) can be reconstructed given an output, y, satisfying

$$||y|| \leq \frac{a(1-K)}{||g_1||_{L^r(0,t_1;\mathbb{R})}||H_0^{-1}||_{L(Y,Z_1)}} \quad \ldots \quad (3.29)$$

Pf. The objective is to apply the contraction mapping theorem (using the second form in which it appears in Appendix 4) to the map Φ defined by (3.28). First we show that Φ is a contraction on B_a.

$$||\Phi z-\Phi\hat{z}||_{L^r(0,t_1;V)} \leq ||g_1||_{L^r(0,t_1;\mathbb{R})} \cdot$$

$$||H_0^{-1}||_{L(Y,Z_1)}||C\int_0^{\cdot} S(\cdot-\tau)(f(z(\tau))-f(\hat{z}(\tau)))d\tau||_Y +$$

$$+ \; ||g_2||_{L^q(0,t_1;\mathbb{R})}||f(z) - f(\hat{z})||_{L^s(0,t_1;Z_2)}$$

using the "convolution argument"

$$\leq (R\|H_0^{-1}\|_{L(Y,Z_1)}\|g_1\|_{L^r(0,t_1;\mathbb{R})} + \|g_2\|_{L^q(0,t_1;\mathbb{R})}) \cdot$$

$$\|f(z) - f(\hat{z})\|_{L^s(0,t_1;Z_2)}$$

by ii)

$$\leq (R\|H_0^{-1}\|_{L(Y,Z_1)}\|g_1\|_{L^r(0,t_1;\mathbb{R})} + \|g_2\|_{L^q(0,t_1;\mathbb{R})}) \cdot$$

$$k(\|z\|_{L^r(0,t_1;V)}, \|\hat{z}\|_{L^r(0,t_1;V)})\|z - \hat{z}\|_{L^r(0,t_1;V)}$$

using iii). Finally, from iv) we conclude

$$\leq K\|z - \hat{z}\|_{L^r(0,t_1;V)} \quad .$$

Hence Φ is a contraction on B_a .

In accord with the second form of the contraction mapping theorem we take $D \equiv B_a$, $w_o = 0$. Then the ball S is given by

$$S \doteq \{z : \|z - S(\cdot)H_0^{-1}y\|_{L^r(0,t_1;V)} \leq \frac{K}{1-K}\|S(\cdot)H_0^{-1}y\|_{L^r(0,t_1;V)}\}$$

Hence for $z_1 \in S$

$$\|z_1\| \leq (1 + \frac{K}{1-K})\|S(\cdot)H_0^{-1}y\|_{L^r(0,t_1;V)}$$

which yields, by i) and (3.29)

$$||z_1|| \leq a$$

and so $S \subset D$.

Thus the hypotheses of the second form of the contraction mapping theorem are satisfied and so Φ has an unique fixed point in B_a . ∎

Exactly as in Corollary 3.14, an additional smoothing hypothesis on the semigroup gives a more regular solution, viz... .

Corollary 3.21

Suppose, in addition to the hypotheses of Theorem 3.20, that $S(t) \in L(Z_2, Z_1)$ for $t > 0$ and satisfies

$$||S(t)z||_{Z_1} \leq g_3(t)||z||_{Z_2} \qquad \forall z \in Z_2$$

where

$$g_3 \in L^{p_3}(0,t_1;\mathbb{R})$$

and

$$p_3 : \frac{1}{p_3} + \frac{1}{s} = 1 \quad .$$

Then the solution proven to exist in $L^r(0,t_1;V)$, by Theorem 3.20, also lies in $C(0,t_1;V)$.

Pf. using the map Φ of (3.28) instead of the mild solution (3.5), one follows the proof of Corollary 3.14 as in Ichikawa-Pritchard [1].

Example 3.22

With reference to Example 3.16 we consider the non-linear wave equation in one dimension and illustrate Theorem 3.20.

$$\begin{aligned} &z_{tt} = z_{xx} + f(z) \\ &z(0,t) = z(1,t) = 0 \qquad \ldots \qquad (3.30) \\ &z(\cdot,0) = z_0(\cdot); z_t(\cdot,0) = z_1(\cdot) \end{aligned}$$

$Az = - z_{xx}$, $D(A) = H^2(0,1) \cap H_0^1(0,1)$

and Q is defined as in Example 2.14.

The semigroup generated by Q has the explicit expression given in Example 2.14.

$$S(t)\begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} \sum_n 2[\langle z_0,\phi_n\rangle \cos n\pi t + \frac{1}{n\pi}\langle z_1,\phi_n\rangle \sin n\pi t]\phi_n \\ \sum_n 2[-n\pi\langle z_0,\phi_n\rangle \sin n\pi t + \langle z_1,\phi_n\rangle \cos n\pi t]\phi_n \end{bmatrix}$$

where $\phi_n = \sin n\pi\xi$.

Suppose we have an observation of the form

$$y(t) = \int_0^1 c(x) z_t(x,t)dx \qquad \ldots \qquad (3.31)$$

where $c(x) = \sum_{n=1}^{\infty} c_n\phi_n(x)$, $c_n = \langle c,\phi_n\rangle$.

Such a $c(x)$ could be used, for example, to model a local spatial average of the time derivative, as an approximation to measurement at a point. Obviously one must require that $c_n \neq 0$, for all n , so as to obtain observability. The time interval of interest is $[0,2]$ i.e. $t_1 = 2$. Suppose now that the output space $\mathcal{Y}$ consists of functions $y(\cdot)$ which can be expressed as

$$y(t) = \sum_{n=1}^{\infty} (a_n \sin n\pi t + b_n \cos n\pi t) \quad \ldots \qquad (3.32)$$

From Definition 2.40, we have that the linear part is continuously initially observable, when $\mathcal{Y}$ is normed by

$$||y||_{\mathcal{Y}}^2 = \sum_{n=1}^{\infty} ((a_n^2 + b_n^2)/c_n^2) \quad \ldots \qquad (3.33)$$

For instance, if $c_n \sim \frac{1}{n}$ then $\mathcal{Y}$ is equivalent to $H^1(0,2)$ (cf Curtain-Pritchard [1]).

Now let $F\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0 \\ w_1^2 \end{bmatrix}$ and make the choices

$$Z_1 = V = D(A^{\frac{1}{2}}) \times L^2(0,1)$$

$$Z_2 = D(A) \times D(A^{\frac{1}{2}})$$

$$p_1 = r = 4 \; ; \; p_2 = q = 4/3 \; ; \; s = 2$$

$$g_1 = g_2 = 1 \quad \text{(constant)}.$$

Then, taking $z = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in L^4(0,2;D(A^{\frac{1}{2}}) \times L^2(0,1))$ and using the fact that (see Appendix 1) $H_0^1(0,1)$, which is $D(A^{\frac{1}{2}})$, is a Banach algebra under pointwise products we have $w_1(t) \in D(A^{\frac{1}{2}})$, $0 \le t \le 2$

and so

$$w_1(\cdot) \in L^4(0,2;D(A^{\frac{1}{2}})) \text{ which implies that}$$

$$w_1^2(\cdot) \in L^2(0,2;D(A^{\frac{1}{2}})) .$$

Hence it makes sense to write

$$||(Fz)(\cdot)||^2_{L^2(0,2;D(A)\times D(A^{\frac{1}{2}}))} = ||w_1^2(\cdot)||^2_{L^2(0,2;D(A^{\frac{1}{2}}))}$$

$$\leq ||z(\cdot)||^4_{L^4(0,2;D(A^{\frac{1}{2}})\times L^2(0,1))}$$

and, similarly, to conclude that iii) of Theorem 3.20 holds with $k(\theta_1,\theta_2) = \gamma(\theta_1+\theta_2)$ for some constant $\gamma > 0$. If we let β_1 denote $||H_0^{-1}||_{L(Y,Z_1)}$ and $\beta_2 = (R\beta_1+1)2\gamma$ then the non-linear observer can be constructed for $||y||_Y \leq a\beta_1^{-1}(1-a\beta_2)$. This expression has its maximum when $a = \frac{1}{2\beta_2}$, i.e. $||y|| \leq \frac{1}{4}\frac{1}{\beta_1\beta_2}$; this corresponds to a contraction constant of $K = \frac{1}{2}$ (we need $\beta_2 a \leq K < 1$). Note that as β_1 decreases (increasing "amplification" in output channel), the ball in the state space reaches a limiting upper size. This is consistent with results concerning the non-existence of solutions to (3.30) for arbitrarily large times and initial data.

The use of Theorem 3.20 depends either on exploiting a smoothing action of the semigroup or, as in Example 3.22, a particular structure possessed by the non-linearity. The formulation has been designed to allow large classes of non-linearities, with the restriction as noted before that only local results are obtained. It is possible to expend much effort on improving the bounds in (3.29). For instance, one could first stabilize, for instance by linear feedback, the linear system so that the semi-group $S(\cdot)$ has bounds

$$||S(t)|| \leq Me^{-\omega t} \qquad \omega > 0 \ .$$

Hence a is increased and so is the size of the ball in Y. It is also possible to use other fixed point theorems in a formulation closely related to that of Theorem 3.20. Here we choose only to present an application of the Schauder fixed point theorem. Applications of the (set-valued) Bohnenblust-Karlin theorem in both its weak and its strong versions will be found in Carmichael-Pritchard-Quinn [1].

Schauder-based result

Recall the map Φ of (3.28) along with the continuous initial observability assumption on the linear part - just as preceded Theorem 3.20.

Theorem 3.23

Let Z_1, Z_2 be Banach spaces; and $p,a,s,s',t_1, R,K \in \mathbb{R}^+$ such that $p \geq s'$, $s \geq 1$, $\frac{1}{s} + \frac{1}{s'} = 1$.

Assume also

i) $\qquad S(t) \in L(Z_2,Z_1) \qquad t > 0$

$$||S(t)z||_{Z_1} \leq g(t)\ ||z||_{Z_2} \quad , \ g \in L^p(0,t_1)$$

ii) $f : C(0,t_1;Z_1) \to L^s(0,t_1;Z_2)$

is continuous and

$$||f(z)||_{L^s(0,t_1;Z_2)} \le \rho(||z||_{C(0,t_1;Z_1)})||z||_{C(0,t_1;Z_1)}$$

where $\rho(\cdot) : \underset{\sim}{\mathbb{R}}^+ \to \underset{\sim}{\mathbb{R}}^+$ is continuous and $\rho(\theta) \to 0$ as $\theta \to 0$.

iii) $$||C\int_0^{\cdot} S(\cdot-\tau)z(\tau)d\tau||_{\mathcal{Y}} \le R||z||_{L^s(0,t_1;Z_2)}$$

iv) $$\int_0^t S(t-\tau)f(\_\_)d\tau : C(0,t_1;Z_1) \to Z_1 \text{ is compact } \forall t \in [0,t_1]$$

the map from $C(0,t_1;Z_1) \to \mathcal{Y}$ defined by

$z(\cdot) \to C\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau$ is compact

v) taking $B_a = \{z \in C(0,t_1;Z_1) : ||z||_{C(0,t_1;Z_1)} \le a\}$

such that

$$(R||H_0^{-1}||_{L(\mathcal{Y},Z_1)}||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} + ||g||_{L^{s'}(0,t_1)}) \sup_{\theta \le a} \rho(\theta) \le K < 1 \ .$$

Then: the state of the system described by (3.24) can be reconstructed, given an output, y , satisfying

$$||y||_{\mathcal{Y}} \le \frac{a(1 - K)}{||H_0^{-1}||_{L(\mathcal{Y},Z_1)}||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)}} \qquad \ldots \quad (3.34)$$

Pf. $(\Phi z)(t) = S(t)H_0^{-1}[y(\cdot) - \int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau] + \int_0^t S(t-\tau)f(z(\tau))d\tau$

gives, by i), iii)

$$||\Phi z||_{C(0,t_1;Z_1)} \le ||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} \; ||H_0^{-1}||_{L(Y,Z_1)} [||y||_Y +$$

$$R||fz||_{L^s(0,t_1;Z_2)}] + ||g||_{L^{s'}(0,t_1;\mathbb{R})} ||fz||_{L^s(0,t_1;Z_2)}$$

and using ii), iv) we have

$$\le ||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} \; ||H_0^{-1}||_{L(Y,Z_1)} ||y||_Y \; +$$

$$[R||H_0^{-1}||_{L(Y,Z_1)} ||S(\cdot)_{\sup_{[0,t_1]} L(Z_1)} + ||g||_{L^{s'}(0,t_1)}].\sup_{\theta\le a} \rho(\theta)a$$

$$\le a(1-K) + Ka \quad , \text{ by v) and (3.34)}$$

$$\le a \; .$$

Hence Φ maps B_a , a closed convex subset of a Banach space, into itself. To show Φ continuous we compute

$$||\Phi(z+h) - \Phi(z)||_{C(0,t_1;Z_1)} \le (R||H_0^{-1}||_{L(Y,Z)} ||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} +$$

$$||g||_{L^{s'}(0,t_1)}) ||f(z+h)-f(z)||_{L^s(0,t_1;Z_2)}$$

and, by ii), Φ is continuous.

Lastly it is required to show that Φ maps B_a into a precompact subset (of B_a) . Here we need iv) and the function-space valued version of the Arzela-Ascoli theorem. Consider first the operator

$$\int_0^t S(t-\tau)f(\_)d\tau : C(0,t;Z_1) \to Z_1$$

$$||\int_0^t S(t-\tau)(fz)(\tau)ds - \int_0^{t_0} S(t_0-\tau)(fz)(\tau)d\tau||_{Z_1} \leq$$

$$||(S(t-t_0) - I)\int_0^{t_0} S(t_0-\tau)(fz)(\tau)d\tau||_{Z_1}$$

$$+ ||\int_{t_0}^t S(t-\tau)(fz)(\tau)d\tau||_{Z_1}$$

$$\leq ||(S(t-t_0) - I)\int_0^{t_0} S(t_0-\tau)(fz)(\tau)d\tau||_{Z_1} +$$

$$||g||_{L^{s'}(t_0,t)} ||(fz)(\cdot)||_{L^s(t_0,t;Z_2)} \quad .$$

The map $E : \mathbb{R} \times Z_1 \to Z_1 : (t,z) \to S(t)z$ is continuous. Let Z_c be a compact subset of Z_1 . E is uniformly continuous on $[0,t_1] \times Z_c$. By iv) the image of B_a under the map $\int_0^{t_0} S(t_0-\tau)f(\_)d\tau$ is compact. Thus

$$||(S(t-t_0) - I)\int_0^{t_0} S(t_0-\tau)f(\_)d\tau||_{Z_1} \to 0 \quad ,$$

as $t \to t_0$, uniformly on B_a . Additionally, $||g||_{L^{s'}(t_0,t)} \to 0$ as $t \to t_0$. Thus we can conclude equicontinuity from the right for

$$\int_0^t S(t-\tau)f(\_)d\tau \quad \text{on } B_a .$$

Now we need to show equicontinuity from the left: take $t > \varepsilon > h > 0$

$$||\int_0^t S(t-\tau)f(z)d\tau = \int_0^{t-h} S(t-h-\tau)f(z)d\tau||_{Z_1} \leq$$

$$||\int_0^{t-\varepsilon} S(t-\tau)f(z)d\tau - \int_0^{t-\varepsilon} S(t-h-\tau)f(z)d\tau||_{Z_1}$$

$$+ ||\int_{t-\varepsilon}^t S(t-\tau)f(z)d\tau||_{Z_1} + ||\int_{t-\varepsilon}^{t-h} S(t-h-\tau)f(z)d\tau||_{Z_1}$$

$$\leq ||(S(\varepsilon)-S(\varepsilon-h))\int_0^{t-\varepsilon} S(t-\varepsilon-\tau)f(z)d\tau||_{Z_1}$$

$$+ 2||g||_{L^{s'}(0,\varepsilon)}||f(z)||_{L^s(0,t_1;Z_2)} .$$

Now let $h \to 0$ and then $\varepsilon \to 0$; from ii) and iii) we have that $\int_0^t S(t-\tau)f(\_)d\tau$ is equicontinuous from the right on B_a .

Now we need to show that $\int_0^t S(t-\tau)f(\_)d\tau$ acting on B_a is uniformly bounded (as a map $B_a \to C(0,t_1;Z_1)$) . Consider

$$||\int_0^{\cdot} S(\cdot-\tau)f(z)d\tau||_{C(0,t_1;Z_1)}$$

$$\leq ||g||_{L^{s'}(0,t_1)}||f(z)||_{L^s(0,t_1;Z_2)}$$

$$\leq ||g||_{L^{s'}(0,t_1)} \sup_{\theta\leq a} \rho(\theta)a .$$

Hence the uniform boundedness; thus we can use the function space valued version of Ascoli-Arzela (see Martin [1]) to conclude that the image of B_a , under $\int_0^t S(t-\tau)f(\_)d\tau$, is compact in $C(0,t_1;Z_1)$.

An entirely analogous argument can be applied to the term Φ_2 where

$$(\Phi_2 z)(t) \to S(t)\ H^{-1}(C\int_0^{\cdot} S(\cdot-\tau)f(z)d\tau) \quad .$$

(We use the second part of iv) and the strong continuity of $S(t)$ exactly as above to conclude equicontinuity of the set $\Phi_2(B_a)$) .

Hence $\Phi(B_a)$ is compact in $C(0,t_1;Z_1)$ and so the conditions of the Schauder theorem are satisfied; thus there is a fixed point of Φ in B_a . ∎

As (a somewhat artificial) example of this result's application : recall Examples 2.14, 2.45 and 3.16 and consider

Example 3.24

$$z_{tt} = z_{xx} + z^2$$

$$z(0,t) = z(1,t) = 0$$

$$C : D(A^{\frac{1}{2}}) \times L^2(0,1) \to L^2(0,1)$$

$$C : \binom{z}{z_t} \to z_t$$ (cf Example 2.45; continuous initial observability by duality)

$$F : \binom{z}{z_t} \to \binom{0}{z^2}$$

$$F : C(0,t_1;\ D(A^{\frac{1}{2}})\times L^2(0,1)) \to L^2(0,t_1;D(A)\times D(A^{\frac{1}{2}}))$$

one has that ii) (for F) is satisfied with $\rho(\theta) = c$ for some constant c (cf Example 3.22, the contraction case). Moreover

$$||\int_0^t S(t-\tau)(Fw(\tau))d\tau||_{D(A)\times D(A^{\frac{1}{2}})} \leq \int_0^t ||z^2||_{D(A^{\frac{1}{2}})} ds$$

$$\leq \int_0^t ||w||^2_{D(A^{\frac{1}{2}})\times L^2(0,1)}$$

gives that $\int_0^t S(t-\tau)(F\, w(\tau))d\tau$ is bounded in $D(A) \times D(A^{\frac{1}{2}})$ if $w(\cdot)$ is bounded in $C(0,t_1;D(A^{\frac{1}{2}})\times L^2(0,1))$. Now $D(A)$ (resp. $D(A^{\frac{1}{2}})$) is compactly embedded in $D(A^{\frac{1}{2}})$ (resp. $L^2(0,1)$). Thus $\int_0^t S(t-\tau)\,(F\, w(\tau))d\tau$ is compact from $C(0,t_1;D(A^{\frac{1}{2}})\times L^2(0,1))$ to $D(A^{\frac{1}{2}})\times L^2(0,1)$. Using the fact that $\int_0^{\cdot} S(\cdot-\tau)(F\, w(\tau))d\tau \in C^1(0,t_1;D(A)\times D(A^{\frac{1}{2}}))$ we have $C\int_0^{\cdot} S(\cdot-\tau)(F\, w(\tau))d\tau \in C^1(0,t_1;D(A^{\frac{1}{2}}))$ and thus, by Appendix 1, compact in $L^2(0,t_1;L^2(0,1))$ as required. Proceeding analogously to Example 3.22 we have that a non-linear observer can be constructed for $||y||_Y \leq a\beta_1^{-1}(1-a\beta_2)$ for appropriate constants β_1,β_2. This gives a fixed point of Φ in $\{z : ||z||_{C(0,t_1;D(A^{\frac{1}{2}})\times L^2(0,1))} \leq a\}$.

In general, the compactness hypotheses of iv) will be satisfied either because a. the operator f is compact, or b. the semigroup $S(\cdot)$ smooths the space Z_1. In both cases one tries to show that $\int_0^t S(t-\tau)f(\_)d\tau$ is bounded from $C(0,t_1;Z_1)$ into Z_0 where Z_0 is compact in Z_1; additionally one needs to show that $C\int_0^{\cdot} S(\cdot - s)f(z)ds$ is compact when considered as a map from $C(0,t_1;Z_1) \to \mathcal{Y}$. Provided $\mathcal{Y}$ can be characterised precisely then one may be able to proceed as before (i.e. showing that image lies in some compactly embedded subspace of $\mathcal{Y}$). Consequently, the known embeddings (Appendix 2) are of great importance in the analysis. In addition, it may be possible, using the theorem of Riesz-Tamarkin (sometimes ascribed to Frechet-Kolmogorov), which is the L^p analogue of Ascoli-Arzela, to consider looking for fixed points in spaces such as $L^p(0,t_1;Z)$.

Some critical comments

The examples above concern hyperbolic partial differential equations. This is no coincidence. For linear parabolic systems (at least without additional manipulation) continuous initial observability rarely obtains. In fact one has ...

Theorem 3.25

Consider the observed system (2.18), (2.19) where A generates a strongly continuous semigroup $S(t)$, $t > 0$, and C is bounded. If we have continuous initial state observability for some $t_1 > 0$ i.e. $\exists\ \gamma \in \mathbb{R}^+$ such that

$$\gamma||CS(\cdot)z_0||_{L^p(0,t_1;Y)} \geq ||z_0||_Z$$

for some p, $1 \leq p < \infty$; and if for each $t \geq 0$ the range of $S(t)$ is dense in Z then $S(t)$ can be extended to a strongly continuous group of bounded operators on $-\infty < t < \infty$.

Pf. See Dolecki-Russell [1].

The results (Theorems 3.2 and 3.23) we have seen demand that H_0 be boundedly invertible on $y(\cdot) - C\int_0^{\cdot} S(\cdot-s)f(z(s))ds$. As the above result makes clear, for a large class of systems (including, at least, linear parabolic ones), this cannot be so when $\mathcal{Y} = L^p(0,t_1;Y)$. The standard procedure (Dolecki-Russell [1] has the most complete discussion of these points) is to restrict attention to the range of H_0 , i.e. take $\mathcal{Y}$ = range(H_0) and then to define a topology on $\mathcal{Y}$ which makes H_0^{-1} continuous. Generally this topology will not be equivalent to the relative topology on range(H_0) inherited from $L^p(0,t_1;Y)$. The most obvious, and robust, way of ensuring that $y(\cdot)-C\int_0^{\cdot} S(\cdot-s)f(z(s))ds \in \mathcal{Y}$ is to demand that both $y(\cdot)$ and $C\int_0^{\cdot} S(\cdot-s)f(z(s))ds$ lie in $\mathcal{Y}$. This is a somewhat stringent requirement; one is in effect asking that both $y(\cdot)$ and $C\int_0^{\cdot} S(\cdot-s)f(z(s))ds$ are given by $CS(\cdot)z_0$ for some z_0's . In order to make sense of the results obtained when the system is initially observable, but not continuously initially observable, with respect to $L^p(0,t_1;Y)$ (and so $\mathcal{Y}$ is taken to be range(H_0)) we must then impose some posteriori verification condition(s) on the fixed point(s) obtained. These conditions will be designed to tell us whether or not the fixed point(s) obtained make sense in terms of the original problem (3.20), (3.21). Such investigations are being performed, but will not be discussed in this thesis. The first part of Chapter 4 presents an alternative way of resolving some of these problems using an optimisation approach. The remainder of this chapter will be devoted to further exploration of the formulation used in Theorems 3.20 and 3.23.

3.3 More applicable non-linear functional analysis

The theorems which have just been applied (contraction, Schauder) are probably the best known of the fixed point results developed by mathematicians over the last 50 or 60 years. Many other results are available and are potentially applicable to the Φ of (3.28). Here, without any claim (or aim) for completeness the use of two other such results is indicated. Further details will be found in Carmichael-Pritchard-Quinn [1] (in addition to Bohuenblust-Karlin) and [2]. As stated above we shall use the formulation of Theorems 3.20 and 3.23, and hence will be subject to the restrictions noted at the end of section 3.2.

Operator splitting

Define $$(\Phi_1(z))(t) = \int_0^t S(t-\tau)f(z(\tau))d\tau \quad \ldots \quad (3.34)$$

and

$$(\Phi_2(z))(t) = S(t)H_0^{-1}(y(\cdot) - C\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau) \quad \ldots \quad (3.35)$$

then, comparing with (3.28),

$$\Phi = \Phi_1 + \Phi_2 \ .$$

Such operator splittings commonly occur in the application of fixed point techniques. A number of theorems have been developed in order to exploit such cases. Here we only use the theorem due to Nussbaum (see Appendix 4). This theorem deals with the sum of a "contraction" and a "compact" operator. In our terms we obtain

Theorem 3.26

Consider the dynamical system described, in mild form, by (3.24) and (3.25); assume that Z_1, Z_2 are Banach spaces $(Z_1 \subset Z_2)$ $p,s \in \mathbb{R}^+ : \frac{1}{p} + \frac{1}{s} = 1$ and

i) the semigroup $S(t)$ generated by A, satisfies

$$S(t) \in L(Z_2, Z_1) \qquad t > 0$$

$$||S(t)z||_{Z_1} \leq g(t)||z||_{Z_2} \quad ; \quad ||g||_{L^p(0,t_1;\mathbb{R})} = c < \infty$$

ii) $R > 0$ is such that

$$||C\int_0^{\cdot} S(\cdot-\tau)z(\tau)d\tau||_{Y} \leq R||z||_{L^s(0,t_1;Z_2)}$$

iii) $f : C(0,t_1;Z_1) \to L^s(0,t_1;Z_2)$

is continuous and satisfies a Lipschitz condition

$$||f(z) - f(\hat{z})||_{L^s(0,t_1;Z_2)} \leq k(||z||,||\hat{z}||)||z - \hat{z}||$$

where the norms on the right hand side are computed in $C(0,t_1;Z_1)$. The function $k(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, symmetric and such that $k(0,0) = 0$,

iv) the map from $C(0,t_1;Z_1) \to Y$ defined by

$z(\cdot) \to C\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau$ is compact.

v) $a \in \mathbb{R}^+$ is chosen so that

$$(Rd + c) \sup_{0 \leq \theta \leq a} k(\theta,0) \leq K < 1$$

where

$$d = ||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} \; ||H_0^{-1}||_{L(Y,Z_1)}$$

and

$$c \sup_{0\le\theta_1,\theta_2\le a} k(\theta_1,\theta_2) \le K < 1 \quad .$$

Then: the state of the system described by (3.24), (3.25) can be reconstructed, given an observation $y(\cdot)$ satisfying

$$||y||_Y \le \frac{a(1-K)}{d}$$

Pf. Consider first

$$z \in B_a = \{z \in C(0,t_1;Z_1): ||z||_{C(0,t_1;Z_1)} \le a\} \quad ; \text{ for such } z \,,$$

$$||\Phi_1 z + \Phi_2 z||_{C(0,t_1;Z_1)} \le ||S(\cdot)||_{\sup_{[0,t_1]} L(Z_1)} \; ||H_0^{-1}||_{L(Y,Z_1)} ||y||_Y \,,$$

$$+ (Rd + c) \sup_{0\le\theta\le a} k(\theta,0)$$

$$\le a(1-K) + Ka = a$$

$\therefore \; \Phi_1 + \Phi_2 : B_a \to B_a$.

The continuity of Φ_1 and Φ_2 follows directly from the continuity of f ; additionally, for $\hat{z}, z \in B_a$,

$$||\Phi_1 z - \Phi_1 \hat{z}||_{C(0,t_1;Z_1)} \le ||g||_{L^p(0,t_1;\mathbb{R})} ||f(z)-f(\hat{z})||_{L^s(0,t_1;Z_2)}$$

$$\le c\, k(||z||,||\hat{z}||)||z - \hat{z}||_{C(0,t_1;Z_1)}$$

$$\le K||z - \hat{z}||_{C(0,t_1;Z_1)}$$

Finally, we need to show that Φ_2 maps B_a into a precompact subset (of B_a) . From condition iv) the image of B_a under $C\int_0^{\cdot} S(\cdot-\tau)f(\_)d\tau$ is precompact in $\mathcal{Y}$. Then by the (strong) continuity of $S(t)$ (and the continuity of H_0^{-1}) we may conclude compactness in $C(0,t_1;Z_1)$.

Thus we have, by Nussbaum a fixed point of $\Phi_1 + \Phi_2$ in B_a . ∎

This theorem may easily be applied to a system such as that of Example 3.24. The critical comments at the end of 3.2 still apply, however. The requirements that H_0^{-1} exists (and is bounded), and that the compactness condition, iv), of the Theorem holds, combine to place severe restriction on the systems which can be studied. Thus, although the proof is much simpler (we do not have to use the pointwise compactness of $\int_0^t S(t-\tau)f(\_)d\tau$) than that of Theorem 3.23 this formulation offers no fundamental improvement. It would be possible to reformulate Theorem 3.26 with Φ_1 as the compact part and Φ_2 as the contraction; this would avoid the need to find a space compactly embedded in $\mathcal{Y}$ but would otherwise have few advantages, hence is not developed here. The operator splitting of Theorem 3.26 will be used again in Chapter 4.

Degree theoretic result

Here we use the degree theoretic formulation (see Appendix 4) of Leray-Schauder's classic paper (Leray-Schauder, [1]). Recall from Appendix 4 the formulation : we consider the equation

$$z - \Phi(z,\lambda) = 0 \qquad \ldots \qquad (3.36)$$

(under a number of assumptions listed in Appendix 4) then, if for some λ_0 we can find all solutions $z - \Phi(z,\lambda_0) = 0$, we may conclude that there is a solution for any λ in some range of interest as long as an associated topological invariant (the Leray-Schauder index) can be calculated at λ_0 and, hence, shown to be non-zero. The aim is to find a λ_0 such that this calculation is particularly easy. In our present case, with an eye to (3.28), we define

$$(\Phi(z,\lambda))(t) = S(t)H_0^{-1}y(\cdot)+\lambda(\int_0^t S(t-\tau)f(z(\tau))d\tau - S(t)H_0^{-1}(C\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau)) \quad \ldots \quad (3.37)$$

Note that $(\Phi(z,0))(t) = S(t)H_0^{-1}(y(\cdot))$ and $\Phi(z,1)$ recovers the Φ of (3.28). Thus in applying the Leray-Schauder result of Appendix 4, we define $\overline{\Omega} = B_a \times M$ where B_a is the ball in $C(0,t_1;Z_1)$ as previously defined, and $M = [0,1]$.

Theorem 3.27

Consider the dynamical system described, in mild form, by (3.24), (3.25); assume that Z_1 , Z_2 are Banach spaces $(Z_1 \subset Z_2)$, $p,s \in \mathbb{R}^+ : \frac{1}{p} + \frac{1}{s} = 1$ and

i) the semigroup $S(\cdot)$ generated by A , satisfies

$$S(t) \in L(Z_2,Z_1) \qquad t > 0$$

$$||S(t)z||_{Z_1} \leq g(t)||z||_{Z_2} \quad ; \quad ||g||_{L^p(0,t_1;\mathbb{R})} = c < \infty$$

ii) $R > 0$ is such that

$$||C\int_0^{\cdot} S(\cdot-\tau)z(\tau)d\tau||_{\mathcal{Y}} \leq R||z||_{L^s(0,t_1;Z_2)}$$

iii) $f : C(0,t_1;Z_1) \to L^s(0,t_1;Z_2)$

is continuous and satisfies a "growth" condition

$$||f(z)||_{L^s(0,t_1;Z_2)} \leq \rho(||z||)||z||$$

where the norms on the right hand side are computed in $C(0,t_1;Z_1)$. The function $\rho(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous and $\rho(\theta) \to 0$ as $\theta \to 0$,

iv) the following compactness conditions are satisfied:

$$\int_0^t S(t-s)f(\_)ds : C(0,t_1;Z_1) \to Z_1 \quad \text{is compact for each} \quad t \in [0,t_1]$$

$$C\int_0^{\cdot} S(\cdot-\tau)f(\_)d\tau : C(0,t_1;Z_1) \to \mathcal{Y} \quad \text{is compact}$$

v) $a \in \mathbb{R}^+$ is chosen so that

$$(Rd + c) \sup_{0\leq\theta\leq a} \rho(\theta) \leq K < 1$$

where

$$d = ||S(\cdot)||_{\sup_{[0,t_1]} \mathcal{L}(Z_1)} ||H_0^{-1}||_{\mathcal{L}(\mathcal{Y},Z_1)}$$

vi) $\partial\Omega$, the boundary of the set Ω does not contain any solution of (3.36).

Then: the state of the system described by (3.24) (3.25) can be reconstructed given an observation, y , satisfying

$$||y||_Y \leq \frac{a}{d} \quad \dots \tag{3.38}$$

Pf. The complete continuity of $\Phi(\cdot,\cdot)$ is shown by proceeding exactly as in Theorem 3.23; hence will not be repeated here. To show uniform continuity with respect to λ, consider

$$\begin{aligned}
&||\Phi(z,\lambda_1) - \Phi(z,\lambda_2)||_{C(0,t_1;Z_1)} \\
&\leq |\lambda_1-\lambda_2| \; ||\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau - S(\cdot)H_0^{-1}C\int_0^{\cdot} S(\cdot-\tau)f(z(\tau))d\tau||_{C(0,t_1;Z_1)} \\
&\leq |\lambda_1-\lambda_2|[||g||_{L^p(0,t_1;\sim)}||f(z)||_{L^s(0,t_1;Z_2)} + Rd||f(z)||_{L^s(0,t_1;Z_2)}] \\
&\leq |\lambda_1-\lambda_2|[Rd + c]\rho(||z||)||z||_{C(0,t_1;Z_1)}.
\end{aligned}$$

and for $z \in B_a$ we have

$$\begin{aligned}
&||\Phi(z,\lambda_1) - \Phi(z,\lambda_2)||_{C(0,t_1;Z_1)} \\
&\leq |\lambda_1-\lambda_2|[Rd + c] \sup_{0\leq\theta\leq a} \rho(\theta)a \\
&\leq Ka|\lambda_1-\lambda_2| \; .
\end{aligned}$$

Thus we obtain uniform continuity with respect to λ.

For $\lambda = 0$ the only solution is $z_L(t) = S(t)\, H_0^{-1}\, y$ and (3.38) ensures that

$$\|z_L\|_{C(0,t_1;Z_1)} \le \|S(\cdot)\|_{\substack{\sup \\ [0,t_1]}} L(Z_1) \|H_0^{-1}\|_{L(Y,Z_1)} \|y\|$$

$$\le d\|y\|_Y$$

$$\le a$$

$\therefore\ z_L \in B_a$.

When $\lambda = 0$, the transformation $z \to z - \Phi(z,\lambda)$ is a translation of the identity and thus has index equal to + 1 .

Thus, provided the "a priori assumption", vi), holds we conclude that $\Phi(\cdot,1)$ has at least one fixed point in B_a . ∎

The most obvious disadvantage of this formulation is the requirement vi). It is this assumption, however, which allows use of a ball in Y having radius a/d ; in general, this is larger than the ball used in previous theorems. In practice a solution technique based on this theorem would attempt to follow the, possibly bifurcating, path of solutions beginning at z_L .

In Leray-Schauder [1] a formula for the (Leray-Schauder) degree of a mapping is developed. In our case the value of the degree is +1 since this is the value at $\lambda = 0$. The formula states that the degree is the sum of the indices of the solutions. Under some supplementary conditions, all the indices must be either +1 or -1 . The calculation of the index at a particular solution will depend on the eigenvalues of the Frechet derivative evaluated at that solution; and, hence, on the behaviour of the linearised system. In Chapter 5 we return to an investigation of the linearised system and implications for questions of uniqueness and algorithms. The directions suggested by direct evaluation of the degree will not be further pursued here; but, it would seem, at least in regard of certain specific problems, such an approach may provide more detailed qualitative information.

Yet another direction which could be pursued (but not here) concerns the introduction of a different "operator-splitting". Suppose we introduce

$$(\Phi(z,\lambda))(t) = S(t)H^{-1}(y(\cdot)) + \int_0^t S(t-s)f(z(s))ds$$
$$- \lambda S(t)H^{-1}\left(C\int_0^{\cdot} S(\cdot-s)f(z(s))ds\right)$$

so that for $\lambda = 0$

$$(\Phi(z,0))(t) = S(t)H^{-1}(y(\cdot)) + \int_0^t S(t-s)f(z(s))ds \ .$$

Then if we show that $z - \Phi(z,0) = 0$ has an unique solution with non-zero index, we may deduce that $z - \Phi(z,1) = 0$ possesses a solution. Typically, this would involve imposing a contraction condition on the operator $\Phi(z,0)$ together with compactness requirements on both $\Phi(z,0)$ and $S(t)H^{-1}\left(C\int_0^{\cdot} S(\cdot-s)f(z(s))ds\right)$. As λ moves from 0 to 1 we follow a continuous path in the space of trajectories, starting at $\lambda = 0$ with the mild solution, (3.24), corresponding to the "linear" estimate of system state i.e. $z(\cdot)$ satisfies

$$z(t) = S(t)z_0 + \int_0^t S(t-s)f(z(s))ds$$

where z_0 is evaluated by solving $z_0 = H^{-1}(y(\cdot))$.

CHAPTER IV : Other topics

Summary

In this chapter we investigate a number of variations on the themes expounded in Chapter III. Specifically we consider a) different ways of constructing the map for whose fixed points we search or b) other control and estimation problems which admit of a fixed-point formulation. In all these cases, as before, the properties of the linear part are exploited in order to create a candidate map for whose fixed points we search. Though the assumptions on the linear part are possibly restrictive the examples of this section by no means exhaust the potential of this approach; see for example Carmichael-Quinn [1].

4.1 Use of pseudo-inverses

Thus far we have demanded (at some cost, see the critical comments of 3.2) continuous initial observability for the linear part and used the existence of a bounded reconstruction operator (H_0^{-1}) to construct the non-linear map Φ . In this section we look at the possibility of using pseudo inverses to provide the bounded reconstruction operator based on the linear part. We shall make use of compactly embedded spaces not only because the linear part gives rise to an ill-posed problem (cf. 2.4) but also because the aim is to apply fixed point theorems which use compactness properties (cf. comments after Ex. 3.24). As has been indicated in previous remarks the obvious application is to those parabolic systems where the linear part is initially observable, but not continuously initially observable.

Consider then the state space of the linear part to be a Hilbert space Z and let Z_1 be another Hilbert space compactly embedded in Z. Recall now Definition 2.49 for ${}_0\tilde{H}^{\dagger}_{\varepsilon,Z_1}$ being the regularised pseudo-inverse with graph norm and thus a map from $Z_1 \times L^2(0,t_1;Y) \to Z_1$, equation (3.28) now becomes

$$(\Phi(z))(t) = S(t)\ {}_0\tilde{H}^{\dagger}_{\varepsilon,Z_1}\left(0, y(\cdot) - \int_0^{\cdot} S(\cdot-s)f(z(s))ds\right) + \int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (4.1)$$

In the manner of (3.34), (3.35) this Φ is split into Φ_1 and Φ_2 defined by

$$(\Phi_1(z))(t) = \int_0^t S(t-s)f(z(s))ds \quad \ldots \quad (4.2)$$

$$(\Phi_2(z))(t) = S(t)\,{}_0H^{-1}_{\varepsilon,Z_1}\left(y(\cdot) - C\int_0^{\cdot} S(\cdot-s)f(z(s))ds\right) \quad \ldots \quad (4.3)$$

Theorem 4.1

Consider the dynamical system described, in mild form, by (3.24) and (3.25) and recast as above; assume that Z_1, H are Hilbert spaces, Z_2 Banach ($Z_1 \subset H \subset Z_2$) with Z_1 compactly embedded in H and H continuously-embedded in Z_2; let $p,s \in \mathbb{R}^+$ and assume that $\frac{1}{p} + \frac{1}{s} = 1$ and, further, that

i) $S(t) \in L(Z_2,H) \qquad t > 0$

$$||S(t)z||_H \le g(t)||z||_{Z_2} \quad ; \quad ||g||_{L^p(0,t_1;\mathbb{R})} = c < \infty$$

ii) $R > 0$ is such that

$$||C\int_0^{\cdot} S(\cdot-\tau)z(\tau)d\tau||_{\mathcal{Y}} \le R||z||_{L^s(0,t_1;Z_2)}$$

where $\mathcal{Y}$ is, as in Chapter 2, now a Hilbert space;
$\mathcal{Y} = L^2(0,t_1;Y)$ for some Y (Hilbert).

iii) $f : C(0,t_1;H) \to L^s(0,t_1;Z_2)$

is continuous and satisfies a Lipschitz condition

$$||f(z) - f(\hat{z})||_{L^s(0,t_1;Z_2)} \le k(||z||,||\hat{z}||)\ ||z-\hat{z}||$$

where the norms on the right hand side are computed in $C(0,t_1;H)$. The function $k(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, symmetric and such that $k(0,0) = 0$.

iv) $a \in \mathbb{R}^+$ is chosen so that

$$(Rd + c) \sup_{0\le\theta\le a} k(\theta,0) \le K < 1$$

where

$$d = ||S(\cdot)||_{\sup_{[0,t_1]} \mathcal{L}(H)} \quad ||{}_0\tilde{H}^{\dagger}_{\varepsilon,Z_1}(0,\cdot)||_{\mathcal{L}(\mathcal{Y},Z_1)}$$

and

$$c \sup_{0\le\theta_1,\theta_2\le a} k(\theta_1,\theta_2) \le K < 1 .$$

Then: the map Φ , defined by (4.1), has a fixed point in the ball $B_a = \{z \in C(0,t_1;H) : ||z||_{C(0,t_1;H)} \leq a\}$ provided that the observation $y(\cdot)$ satisfies

$$||y||_Y \leq \frac{a(1-K)}{d} .$$

Pf. Consider first $z \in B_a$; for such a z we have by the hypotheses above

$$||\Phi_1 z+\Phi_2 z||_{C(0,t_1;H)} \leq ||S(\cdot)||_{\sup_{[0,t_1]} L(H)} \; ||{}_0\tilde{H}_{\varepsilon,Z_1}(0.\cdot)||_{L(Y,Z_1)}||y||_Y +$$

$$+ (Rd+c) \sup_{0\leq\theta\leq a} k(\theta,0) \leq a(1-K) + Ka = a$$

$\therefore \Phi_1 + \Phi_2 : B_a \to B_a$.

The continuity of Φ_1 and Φ_2 follows from that of f ; additionally

$$||\Phi_1 z-\Phi_1\hat{z}||_{C(0,t_1;H)} \leq K||z-\hat{z}||_{C(0,t_1;H)}$$

exactly as in the proof of Theorem 3.26.

Finally we need to show that Φ_2 maps B_a into a precompact subset of B_a . From the definition of ${}_0\tilde{H}_{\varepsilon,Z_1}$, it maps B_a into a precompact subset of H . Then by the (strong) continuity of $S(t)$ we may conclude compactness in $C(0,t_1;H)$.

Thus we have, by Nussbaum a fixed point of $\Phi_1 + \Phi_2$ in B_a . ∎

Consider the example

<u>Example 4.2</u> (cf Example 2.43)

$$z_t = z_{xx} - zz_x \quad ; \quad z(0,t) = z(1,t) = 0$$

$$y(t) = \int_0^1 c(x)z(x,t)dx \qquad (y(\cdot) \in L^2(0,t_1;\mathbb{R})) \quad .$$

Then, as in Example 2.43, the linear part is initially observable if

$$\int_0^1 c(x) \sin n\pi x \, dx \neq 0 \qquad \forall n \; .$$

Even if this holds, it is not, however, continuously so. Here we make the slightly artifical assumption that

$$Z_1 = D(A) \; ; \; H = H_0^1(0,1) \; ; \; Z_2 = L^2(0,1)$$

$f : \psi \to -\psi\psi_x : H \to Z_2$ is proven as follows -

recall that for $\phi \in H_0^1$, $\phi(\cdot) = \int_0^{\cdot} \psi'(x)dx$. (ψ' used here instead of ψ_x for convenience); so ψ is absolutely continuous and $\sup_x |\psi(x)| \le ||\psi||_{H_0^1}$; thus

$$||\psi_1\psi_1' - \psi_2\psi_2'||_{Z_2} \le ||\psi_1(\psi_1'-\psi_2')||_{Z_2} + ||(\psi_1-\psi_2)\psi_2'||_{Z_2}$$

$$\le ||\psi_1||_H \; ||\psi_1-\psi_2||_H + ||\psi_1-\psi_2||_H \; ||\psi_2||_H$$

$$\le (||\psi_1||_H + ||\psi_2||_H) \; ||\psi_1-\psi_2||_H \; .$$

We take $p = 2-\varepsilon$ $\varepsilon \in]0,1[$ (this comes from the definition of the norm on $||S(t)z||$) and s correspondingly. $D(A)$ is compactly embedded in H which is continuously embedded in Z . Thus with appropriate $a,c,d,K,R \in \mathbb{R}^+$ we may satisfy the hypotheses of Theorem 4.1; and hence conclude the existence of a fixed point for the map Φ . This point will be a "system trajectory" lying in $C(0,t_1;H)$.

As in Chapter 2 the behaviour of the fixed point as $\varepsilon \to 0^+$ is a natural question to study; as are continuity properties with respect to the data $y(\cdot)$. Notice that since ${}_0\tilde{H}^{+}_{\varepsilon,Z_1}$ recovers a best approximation, the fixed point can now only be regarded as a "consistent" state trajectory. The relation between this result and those obtained by more direct attacks on the non-linear optimisation problem is also worthy of investigation. The result does provide, however, an approach to state reconstruction which is appropriate to a class of non-linear parabolic problems; which class cannot sensibly be handled by the methods of Chapter 3.

Another choice of pseudo-inverse

As was indicated in Chapter 2 one can think of many linear problems where the pseudo-inverse provides a useful notion of solution. Now, if semi-linear terms are added it may be possible to use the pseudo-inverse derived from the linear part to create a map whose fixed points provide a notion of state reconstruction appropriate to the non-linear problem. Here we consider one example of this procedure.

Recall the map T of (2.37) in the case where, for a linear parabolic equation, both the semigroup and the Lions' formulations apply. Recall also the "regularised" operator $\tilde{T}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1}$, defined for Theorem 2.51 (and the other notations used there: viz. Z_1 compactly emebdded in H ; Z_1, H being Hilbert; $C : L^2(0,t_1;H) \to Y)$. The notation $P_{W_{Z_1}}$ will be used to denote the projection of the product space $W_{Z_1}(0,t_1) \times Z_1$ onto its first factor $W_{Z_1}(0,t_1)$, such that $P_{W_{Z_1}}(z(\cdot),z_0) = z(\cdot)$. The semi-linear problem (3.24), (3.25) then naturally gives rise to consideration of the following map Φ

$$\Phi z = P_{W_{Z_1}} \tilde{T}^{\dagger}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1}(0,0,\int_0^{\cdot} S(\cdot-s)f(z(s))ds,y(\cdot)) \quad \ldots \quad (4.4)$$

(recall that

$$\tilde{T}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1} : W_{Z_1}(0,t_1)\times Z_1 \to W_{Z_1}(0,t_1)\times Z_1 \times L^2(0,t_1;H)\times Y$$

as in Chapter 2).

This formula arises as follows: we would like to use the map T of (2.37) to find a solution of

$$T(z,z_0) = (\int_0^{\cdot} S(\cdot-s)f(z(s))ds,y(\cdot)) \quad \ldots \quad (4.5)$$

where $\int_0^{\cdot} S(\cdot-s)f(z(s))ds \in L^2(0,t_1;H)$, $y(\cdot) \in Y$. Being aware, however, of the scarcity of continuous initial observability (at least among parabolic equations) and of the smoothing properties of the semigroup action we are driven to consider the regularised problem

$$\tilde{T}_{\varepsilon,L^2(0,t_1;H)\times H}(z,z_0) = (0,0,\int_0^{\cdot} S(\cdot-s)f(z(s))ds, y(\cdot)) \quad \ldots \quad (4.6)$$

From both a practical and a theoretical viewpoint it is desirable to have some continuity properties with respect to the right hand side of (4.6). Thus we look to

$$\tilde{T}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1}(z,z_0) = (0,0,\int_0^{\cdot} S(\cdot-s)f(z(s))ds,\ y(\cdot)) \quad . . . \qquad (4.7)$$

Under appropriate conditions this can be made consistent with the mild and Lions representations (e.g. the commentary following Theorem 2.51).

As we know that $W_{Z_1}(0,t_1)$ is compactly embedded in $L^2(0,t_1;H)$ we shall aim in the following theorem to use Schauder applied to Φ, thought of as a map from $L^2(0,t_1;H) \to L^2(0,t_1;H)$.

Theorem 4.3

Consider the dynamical system described, in mild form, by (3.24) and (3.25) and recast as above; assume that Z_1, H are Hilbert spaces, Z_2 Banach ($Z_1 \subset H \subset Z_2$) with Z_1 compactly embedded in H and H continuously embedded in Z_2 ; let the constant for the first embedding be given by $e_1 \in \mathbb{R}^+$, i.e.

$$||z||_H \leq e_1 ||z||_{Z_1} \quad .$$

Take $\varepsilon > 0$ and $p,s \in \mathbb{R}^+$ such that $p \geq 1$, $s \geq 1$ and $\frac{1}{p} + \frac{1}{s} = \frac{3}{2}$. Further, assume that

i) $S(t) \in L(Z_2,H) \qquad t > 0$

$$||S(t)z||_H \leq g(t)||z||_{Z_2} \quad ; \quad ||g||_{L^p(0,t_1;\mathbb{R})} = c < \infty$$

ii) $f : L^2(0,t_1;H) \to L^s(0,t_1;Z_2)$

is continuous and such that

$$||f(z)||_{L^s(0,t_1;Z_2)} \le \rho(||z||_{L^2(0,t_1;H)})\, ||z||_{L^2(0,t_1;H)}$$

where $\rho(\cdot) : \underset{\sim}{\mathbb{R}}^+ \to \underset{\sim}{\mathbb{R}}^+$ is continuous and $\rho(\theta) \to 0$ as $\theta \to 0$.

iii) $a \in \underset{\sim}{\mathbb{R}}^+$ is chosen so that

$$c\, d \sup_{\theta \le a} \rho(\theta) \le K < 1$$

where

$$d = e_1\, ||P_{W_{Z_1}} \tilde{T}^{\dagger}_{\varepsilon, W_{Z_1}(0,t_1)\times Z_1} (0,0,..,.)||$$

(the norm is taken in $\mathcal{L}(L^2(0,t_1;H)\times \mathcal{Y}\ ,\ W_{Z_1}(0,t_1)))$.

Then: the map Φ, defined by (4.4), has a fixed point in the ball $B_a = \{z \in L^2(0,t_1;H) : ||z||_{L^2(0,t_1;H)} \le a\}$ provided that the observation $y(\cdot)$ satisfies

$$||y||_{\mathcal{Y}} \le \frac{a(1-K)}{d}$$

Pf. Consider first $z \in B_a$; for such a z we have

$$||\Phi z||_{L^2(0,t_1;H)} \le e_1 ||\Phi z||_{W_{Z_1}(0,t_1)}$$

$$\le e_1 ||P_{W_{Z_1}} \tilde{T}^{\dagger}_{\varepsilon, W_{Z_1}(0,t_1)\times Z_1}||\cdot(||\int_0^{\cdot} S(\cdot-\tau) f(z(\tau))d\tau||_{L^2(0,t_1;H)} + ||y||_{\mathcal{Y}})$$

and using the "convolution property" as before we have

$$\leq d\ c \sup_{\theta \leq a} \rho(\theta)\ a + a(1-K)$$

$$\leq a \quad .$$

Thus $\Phi : B_a \to B_a$. The continuity of Φ is straightforward.

As $\Phi(B_a) \subset W_{Z_1}(0,t_1)$ and this latter space is compactly embedded in $L^2(0,t_1;H)$ we have that $\overline{\Phi(B_a)}$ is a compact subset of $L^2(0,t_1;H)$ and thus of B_a . Thus we have, by Schauder, the existence of a fixed point for Φ in B_a . ∎

In the following we use the notation, for $z \in L^2(0,1)$, $|z|^2 = \int_0^1 (z(x))^2 dx$.

Example 4.4 (c. Example 2.43)

$$z_t = z_{xx} - |z|\ z \quad ; \quad z(0,t) = z(1,t) = 0$$

$$y(t) = \int_0^1 c(x) z(x,t)\ dx \quad .$$

Then the linear part is initially observable if

$$\int_0^1 c(x) \sin n\pi x\ dx \neq 0 \quad \forall n$$

even assuming this to hold it is not, however, continuously so.

Take $Z_1 = H_0^1(0,1)$; $Z_2 = H = L^2(0,1)$.

It is clear that

$$f(\cdot) : L^2(0,t_1;H) \to L^1(0,t_1;Z_2) \quad .$$

Therefore we take $s = 1$, $p = 2$ and

$$\rho(r) = r \quad .$$

It is worth noting here that by Lemma 2.18 (or results of Lions, [1])

$$z_t = z_{xx} + g \quad ; \quad z(0,t) = z(1,t) = 0$$

has a solution in $C(0,t_1;H)$ for $g \in L^1(0,t_1;Z_2)$.
Interpolation techniques are used in Lions [4] to consider a wider class of forcing terms g; e.g. any g in $L^{q(\beta)}(0,t_1;H^{-\beta}(0,1))$ where $\frac{1}{q(\beta)} = 1 - \frac{\beta}{2}$, $0 < \beta < 1$, will still give a solution z in $C(0,t_1;L^2(0,1))$. Such results are used to consider other non-linearities, for instance, $|z|^2 z$.

To satisfy condition i) estimates on $||S(\cdot)||$ are needed. Condition iii), however, required estimate of an operator norm. How best to perform this estimation is not clear. When in finite dimensional state spaces one can construct (using the normal equations) explicit representations of the $\tilde{T}^{\dagger}$.... operator; and use these to state hypotheses ensuring the required norm bounds.

Nonetheless, from a computational viewpoint, this approach has some attractions. The numerical solution of linear least squares problems such as is expressed by $\tilde{T}_{\varepsilon,W_{Z_1}(0,t_1)\times Z_1}$, has been much studied.
In any iterative method based on this formulation the use of compactly embedded spaces will provide at each step desirable continuity properties with respect to the data. As in the comments following Example 4.2, however, the behaviour of the fixed point as $\varepsilon \to 0^+$ is a natural question to study.

Another, and more significant, difficulty in interpretation is the relation between the fixed point obtained and the solution of the (deterministic) non-linear optimisation problem posed as "find $z(\cdot)$ such that

$$z(t) = S(t)z_0 + \int_0^t S(\cdot - s)f(z(s))ds$$

which minimizes

$$\int_0^{t_1} \langle z(\tau), z(\tau)\rangle d\tau + \int_0^{t_1} \langle Cz(\tau)-y(\tau), Cz(\tau)-y(\tau)\rangle d\tau."$$

For the solution of this problem may be available by other means (e.g. maximum principle); examination of the simplest cases shows this solution to be different from that obtained using the fixed point approach. This is perhaps not surprising when one considers that the non-linear optimisation procedure makes essential use of gradient information about the non-linearity; whereas the fixed point approach never calculates the gradient - one might regard it as possibly providing the "best approximation without differentiation". The following (formal) elaboration concerns this point.

Consider that

$$J_{t_1} = \tfrac{1}{2}\langle z(0)-\overline{z}_0, P_0^{-1}(z(0)-\overline{z}_0)\rangle + \tfrac{1}{2}\int_0^{t_1} \langle w(s), Q^{-1}(s)w(s)\rangle ds$$
$$+ \tfrac{1}{2}\int_0^{t_1} \langle y(s)-Cz(z), R^{-1}(y(s)-Cz(s))\rangle ds \quad \ldots \quad (4.8)$$

is to be minimised with respect to $z(t)$, $w(t)$ $0 \le t \le t_1$, subject to the constraint

$$\frac{dz}{dt} = Az + f(z) + w \qquad 0 \le t \le t_1 \qquad \ldots \qquad (4.9)$$

In view of the constraint we minimise J_{t_1} w.r.t. $z(0)$ and $w(\cdot)$; the best estimate for $z(\cdot)$ is then determined. Proceeding in the usual fashion, we form the augmented cost functional

$$J^a_{t_1} = J_{t_1} + \int_0^{t_1} \langle\lambda(s),(\dot{z}-Az-f(z)-w)(s)\rangle ds \qquad \ldots \qquad (4.10)$$

and compute the first variation of $J^a_{t_1}$. Here the exposition bifurcates for we shall a) keep the nonlinearity $f(z)$ or b) replace it by a (known) perturbation $g(\cdot)$: $(P_0,Q,R$ +ve definite, self-adjoint);

a) gives

$$\delta J^a_{t_1} = \langle P_0^{-1}(z(0)-\bar{z}_0)-\lambda(0),\delta z(0)\rangle + \langle\lambda(t_1),\delta z(t_1)\rangle$$
$$+ \int_0^{t_1} \{\langle(R^{-1}C)^*(y(s)-Cz(s)) + \dot{\lambda}(s) + A^*\lambda + (df|_{z(\cdot)})^*\lambda,\delta z(s)\rangle$$
$$+ \langle Q^{-1} w(s)-\lambda(s),\delta w(s)\rangle\}ds \qquad \ldots \qquad (4.11)$$

Thus necessary conditions for $\delta J^a_{t_1} = 0$ are given by

$$P_0^{-1}(z(0)-\bar{z}_0) - \lambda(0) = 0 \qquad \ldots \qquad (4.12)$$

$$\lambda(t_1) = 0 \qquad \ldots \qquad (4.13)$$

$$\frac{d\lambda}{dt} = -(A + df|_z)^*\lambda - C^*R^{-1}(y(t) - Cz(t)) \qquad \ldots \qquad (4.14)$$

$$Q^{-1}w(s) - \lambda(s) = 0 \qquad \ldots \qquad (4.15)$$

Eliminating $w(\cdot)$ from (4.9) using (4.15) we obtain

$$\frac{d\lambda}{dt} = -\left(A + df\big|_z\right)^* \lambda - C^* R^{-1}(y(t) - Cz(t))$$

$$\frac{dz}{dt} = Q\lambda(t) + Az(t) + fz(t) \qquad \ldots \qquad (4.16)$$

$$z(0) = \bar{z}_0 + P_0\lambda(0)$$

$$\lambda(t_1) = 0 \; .$$

Then (4.16) is, in effect, the "optimality system" of Lions;

b) gives an optimality system (by almost identical calculations)

$$\frac{d\lambda}{dt} = -A^*\lambda - C^* R^{-1}(y(t) - Cz(t))$$

$$\frac{dz}{dt} = Q\lambda(t) + Az(t) + g(t) \qquad \ldots \qquad (4.17)$$

$$z(0) = \bar{z}_0 + P_0\lambda(0)$$

$$\lambda(t_1) = 0 \; .$$

When we substitute $f(z)$ for g in (4.17) we obtain a fixed point formulation of the type discussed in the preceding paragraphs. It is clear that, in general, the resulting fixed point will not be a solution of (4.16) unless $df\big|_z = 0$. Thus we cannot expect to attain the non-linear optimum, using the fixed point approach.

Another question of interest is suggested by (4.16): given that the essence of the fixed point approach is to reach the solution of a non-linear problem via a sequence of linear ones, can we formulate (4.16) in this way? A possible answer is indicated by consideration of the iteration defined by

$$\frac{d\lambda_n}{dt} = -(A + df|_{z_{n-1}})^* \lambda_n - C^* R^{-1}(y(t) - Cz(t))$$

$$\frac{dz_n}{dt} = Q\lambda_{n-1} + (A + df|_{z_{n-1}})z_n + (f - df|_{z_{n-1}})(z_{n-1})$$

$$z_n(0) = \bar{z}_0 + P_0\lambda_n(0) \qquad \ldots \quad (4.18)$$

$$\lambda_n(t_1) = 0 \quad .$$

This iteration corresponds to solving the reconstruction problem, i.e. minimising cost functional (4.8), for the linear system

$$\dot{z}_n = (A + df|_{z_{n-1}})z_n + (f - df|_{z_{n-1}})(z_{n-1}) \quad .$$

The proposed scheme is closely related, it seems, to methods of quasi-linearisation, such as are studied in Falb-Jong, [1], and is further investigated in Carmichael-Quinn [1]. As will be seen in Chapter 5, application of Newton's method to the fixed point problem for the Φ of (3.28) yields a scheme with a similar structure.

4.2 State and parameter estimation

Suppose we are given a system of the form

$$\dot{z} = f(z,\alpha) \qquad z(0) = z_0$$

$$y = h(z)$$

and that we know neither the state $z(\cdot)$ nor the parameters α. Both these have to be recovered from the output $y(\cdot)$. Now make some initial guess $(\bar{z}(\cdot),\bar{\alpha})$ and construct a local approximation about this guess. With simplification (largely, as indicated in the introduction, for ease of exposition) we shall assume that the local approximation gives equations of the form

$$\begin{aligned} \dot{z} &= Az + A_1\alpha + f(z,\alpha) \qquad z(0) = z_0 \\ y &= Cz \end{aligned} \qquad \ldots \qquad (4.19)$$

The parameters α will be assumed to be constants and hence we may, at least formally, describe the system by

$$\begin{bmatrix} \dot{z} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} A & A_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ \alpha \end{bmatrix} + \begin{bmatrix} f(z,\alpha) \\ 0 \end{bmatrix}, \quad \begin{bmatrix} z \\ \alpha \end{bmatrix}(0) = \begin{bmatrix} z_0 \\ \alpha \end{bmatrix} \qquad \ldots \qquad (4.20)$$

$$y = [C, \; 0] \begin{bmatrix} z \\ \alpha \end{bmatrix} \qquad \ldots \qquad (4.21)$$

The problem of joint state and parameter estimation has thus been recast as a "semi-linear estimation problem" of the type dealt with in this thesis. We may hope to apply, under appropriate conditions, the fixed point results of the preceding sections. We may thus expect to produce algorithms (with some associated convergence analysis) for state and parameter estimation. This in itself is sufficiently unusual to merit attention - see for instance the discussion of Chavent [1] for an account of the difficulties involved in arriving at a definition of identifiability which is both analytically applicable and practically productive. Suppose we consider our parameter to be constant in $\underset{\sim}{\mathbb{R}}^p$ then we must study the injectivity of the map from the space of initial states $\times \underset{\sim}{\mathbb{R}}^p$ to the space of outputs arising from the linear system

$$\begin{bmatrix} \dot{z} \\ \alpha \end{bmatrix} = \begin{bmatrix} A & A_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z \\ \alpha \end{bmatrix} ; \begin{bmatrix} z \\ \alpha \end{bmatrix}(0) = \begin{bmatrix} z_0 \\ \alpha \end{bmatrix} \quad \ldots \quad (4.22)$$

$$y = [C \quad 0] \begin{bmatrix} z \\ \alpha \end{bmatrix} \quad \ldots \quad (4.23)$$

A useful criterion for ensuring this property is .... .

Lemma 4.5

i) Suppose that the system

$$\begin{aligned} \dot{z} &= Az \qquad z(0) = z_0 \\ y &= Cz \end{aligned} \quad \ldots \quad (4.24)$$

is continuously initially observable on $[0,t_1]$;

ii) the map from $D(A) \times \underset{\sim}{\mathbb{R}}^p \to Z \times Y$ defined by

$\begin{pmatrix} z_0 \\ \alpha \end{pmatrix} \to \begin{pmatrix} Az_0 + A_1\alpha \\ Cz_0 \end{pmatrix}$ is injective;

Then: a. assumptions i) and ii) together imply that the augmented system is continuously initially observable (i.e. the map from $\binom{z_0}{\alpha} \rightarrow y(\cdot)$ defined by (4.22), (4.23) is injective and has closed range);

b. the augmented system being continuously initially observable implies i) and ii).

Pf. a. Consider the mild expression of (4.22), (4.23) viz.

$$y(t) = CS(t)z_0 + C\int_0^t S(t-s)A_1\alpha \, ds \qquad \ldots \qquad (4.25)$$

and suppose that this output is identically zero on $[0,t_1]$. Then $Cz_0 = 0$ by evaluation at $t = 0$. Forming $\frac{y(t+h) - y(t)}{h}$, by use of (4.25), we have that $CS(t)\left(\frac{S(h) - I}{h}\right)z_0 \rightarrow$ a limit as $h \rightarrow 0$. Hence by i) $\frac{S(h) - I}{h}z_0$ has a limit as $h \rightarrow 0$ and so $z_0 \in D(A)$, by definition. Thus (using Theorem 2.5) we may differentiate (4.25) to obtain

$$CS(t)(Az_0 + A_1\alpha) = 0 \quad \text{on } [0,t_1]$$

(again using Theorem 2.5; and, on the integral term, a change of variables, $u = t-s$) and as the system (4.24) is initially observable we have that $Az_0 + A_1\alpha = 0$. By ii) $Cz_0 = 0 = Az_0 + A_1\alpha$ can only occur when $z_0 = 0$, $\alpha = 0$. Thus the map (42.5) is injective and so the augmented system is initially observable.

To show *continuous* initial observability (required if the "inversion" procedure, described in Chapter III, for creating (3.28) is to work) all we need note is that the right hand term of (4.25) is defined on a finite dimensional space. Hence by Example 2.31, c., and the results preceding Theorem 2.48 we may conclude that the augmented system is continously initially observable.

b. Again by Example 2.31, c., continuous initial observability of the augmented system implies that of (4.24). If the injectivity assumption on $\begin{pmatrix} A & A_1 \\ C & 0 \end{pmatrix}$ does not hold then by (4.25) (and its time derivative) there will exist a non trivial (z_0,α) such that the corresponding $y(0) = 0$ and $\frac{dy}{dt} = 0$ i.e. output is zero. This contradicts continuous initial observability for the augmented system. ∎

The condition

$$\ker\begin{pmatrix} A & A_1 \\ C & 0 \end{pmatrix} = \{0\} \qquad \ldots \qquad (4.26)$$

though simple, is slightly novel and has some relevance to procedures for joint state and parameter estimation using Kalman filters (see Jazwinski, [1]). The argument simplifies when one is using a finite dimensional state space. This is the case in Example 4.6. In the case that $S(\cdot)$ is a group and (4.24) is continuously initially observable we have another approach. For then in case that (z_0,α) gives $y(\cdot) \equiv 0$ we obtain at each instant t'

$$z_0 = -\int_0^{t'} S(-s)A_1\alpha \, ds$$

which lies in $D(A)$ (using Theorem 2.5 and the fact that continuous initial observability implies continuous final observability; see Curtain-Pritchard [1], p.70).

Having established continuous initial observability for the augmented system we are free to construct a Φ (for the "augmented state trajectory") exactly as in (3.28) but now in terms of the output operator and semigroup action of (4.22), (4.23). Then we may apply Theorem 3.20 (or even Theorem 3.23) to show that in some ball in the augmented state space Φ has a fixed point. Moreover, if one uses Theorem 3.20, a successive approximation procedure will converge to this fixed point. Rather than repeat the formulation of Theorems 3.20, 3.23 we illustrate the approach with two examples.

<u>Example 4.6</u>

Consider the finite dimensional system $(\underset{\sim}{x} \in \underset{\sim}{\mathbb{R}}^2)$

$$\dot{\underset{\sim}{x}} = \begin{pmatrix} 0 & 1 \\ 1 & \alpha_1 \end{pmatrix} \underset{\sim}{x}$$

$$y(t) = (1 \quad 0)\, \underset{\sim}{x}$$

where $\alpha \in \underset{\sim}{\mathbb{R}}^1$ is a constant. Assume initial guesses, the constants $\overline{x}, \overline{\alpha}$, for state trajectory and parameter respectively; then set $\overline{x} + z = x$, $\overline{\alpha} + \alpha = \alpha_1$ to obtain for $\underline{z} \in \underset{\sim}{\mathbb{R}}^2$ $(\underset{\sim}{z} = (z_1, z_2)^T)$

$$\dot{\underset{\sim}{z}} = \begin{pmatrix} 0 & 1 \\ 1 & \overline{\alpha} \end{pmatrix} z + \begin{pmatrix} 0 \\ \alpha\overline{x} \end{pmatrix} + \begin{pmatrix} 0 \\ \overline{\alpha x} \end{pmatrix} + \begin{pmatrix} 0 \\ \alpha z_2 \end{pmatrix}$$

$$y = (1 \quad 0)\, z + (1 \quad 0)\, \overline{x}$$

For simplicity, take $\overline{\alpha} = 0$; thus we obtain a system

$$\dot{\underset{\sim}{z}} = A \underset{\sim}{z} + A_1 \alpha + f(z, \alpha)$$

$$y = Cz + h$$

where h is a known function and (A,C) observable; the preceding treatment needs only minor modifications in order to account for the presence of $h(\cdot)$. Now $z = (z_1,z_2) \in \ker C \Rightarrow z_1 = 0$ and therefore from (4.26) we obtain $\alpha\overline{x} = 0$ and $z_2 = 0$; if we take $\overline{x} \neq 0$ then we have

$$\ker\begin{pmatrix} A & A_1 \\ C & 0 \end{pmatrix} = \{0\} .$$

Hence we may apply our fixed point results; in Theorem 3.20, for example, the non-linearity satisfies the contraction condition with $k(\theta_1,\theta_2) = c(\theta_1+\theta_2)$ for some constant $c \in \mathbb{R}^+$.

<u>Example 4.7</u> (cf. Examples 3.16, 322)

Consider the observed wave equation in one dimension

$$w_{tt} = w_{xx} + \alpha_1 w \quad \ldots \quad (4.27)$$

$$w(0,t) = w(1,t) = 0 \quad \ldots \quad (4.28)$$

$$y(t) = \int_0^1 c(x)w(x,t)dx \quad \ldots \quad (4.29)$$

where $\alpha \in \mathbb{R}$ is an unknown parameter. Assume initial guesses $\overline{z},\overline{\alpha}$ for state and parameter respectively. Let $\overline{z}$ be independent of time and satisfy the boundary condition of (4.28). Then set

$$z + \overline{z} = w , \quad \alpha + \overline{\alpha} = \alpha_1$$

to obtain

$$z_{tt} = z_{xx} + \overline{\alpha}z + \overline{z} + \overline{z}_{xx} + \overline{\alpha}\,\overline{z} + \alpha z \quad \ldots \quad (4.30)$$

$$z(0,t) = z(1,t) = 0 \quad \ldots \qquad (4.31)$$

$$y(t) = \langle c,z\rangle_{L^2(0,1)} + \langle c,\bar{z}\rangle_{L^2(0,1)} \quad \ldots \qquad (4.32)$$

Assume for simplicity that $\bar{\alpha} = 0$, and so using the notation familiar from other examples with the wave equation we form the augmented state $(z,z_t)^T$ and obtain

$$\begin{pmatrix} z \\ z_t \end{pmatrix}_t = \begin{pmatrix} 0 & I \\ \frac{\partial^2}{\partial x^2} & 0 \end{pmatrix}\begin{pmatrix} z \\ z_t \end{pmatrix} + \begin{pmatrix} 0 \\ \bar{z} \end{pmatrix}\alpha + \begin{pmatrix} 0 \\ \bar{z}_{xx} \end{pmatrix} + \begin{pmatrix} 0 \\ \alpha z \end{pmatrix}$$

Hence we have an equation in the form

$$\dot{Z} = AZ + A_1\alpha + f(Z,\alpha) + g$$

$$\hat{y} = CZ$$

where $\hat{y}(\cdot) = y(\cdot) - \langle c,\bar{z}\rangle_{L^2(0,1)}$.

The known function $\langle c,\bar{z}\rangle_{L^2(0,1)}$ causes no difficulties.

Recall Example 2.14 and the eigenfunctions ϕ_n , defined there. Set $c_n = \langle c,\phi_n\rangle$, and assume $c_n \neq 0 \ \forall n$. Taking $\mathcal{Y}$ to be the space of functions of the form $y(t) = \sum_{n=1}^{\infty} (a_n \cos n\pi t + b_n \sin n\pi t)$ normed by $||.|| = \sum_{n=1}^{\infty} n^2\pi^2 \frac{a_n^2 + b_n^2}{c_n^2}$, the pair (A,C) is continuously initially observable.

By condition (4.26) the linear part of the augmented system is continuously initially observable if the conditions

$$z_t = 0 \qquad \ldots \qquad (4.33)$$

$$\frac{\partial^2 z}{\partial x^2} + \alpha\bar{z} = 0 \qquad \ldots \qquad (4.34)$$

$$\langle c,z\rangle = 0 \qquad \ldots \qquad (4.35)$$

imply that $z = 0$, $z_t = 0$, $\alpha = 0$. Setting

$$z(x,\cdot) = \sum_{n=1}^{\infty} z_n(\cdot)\phi_n(x) \ , \quad \bar{z}_n = \langle\bar{z},\phi_n\rangle$$

(4.34) becomes

$$-\,n^2\pi^2 z_n + \alpha\bar{z}_n = 0 \qquad \forall n \qquad \ldots \qquad (4.36)$$

Substituting (4.36) in (4.35) gives

$$\alpha \sum_{n=1}^{\infty} \frac{c_n\,\bar{z}_n}{n^2\pi^2} = 0 \ .$$

Hence if one assumes $c,\bar{z}$ are such that $\sum_{n=1}^{\infty} \frac{c_n\,\bar{z}_n}{n^2\pi^2} \neq 0$

then the linear part of the augmented system is continuously initially observable. Clearly the non-linearity is a local contraction on $H_0^1(0,1) \times L^2(0,1) \times \mathbb{R}$ into $L^2(0,1)$. Then one can conclude that, subject to conditions (as in Theorem 3.20) on the operators and the output, iteration of a map Φ (based on (3.28), but taking account of g and $\hat{y}$) will determine both the state z and the parameter α .

The condition imposed on the linear part is designed to ensure the injectivity of the "initial state $\times$ parameter output" map for the linear system resulting from linearisation about some nominal state trajectory and parameter value. This is identical with the requirement imposed in some other works which attempt to provide a rational basis for identification algorithms (again, see Chavent [1]). In some such work the initial state is assumed known (in Chavent's case the problem concerns identification of parameters in a wave equation given the observed response to a seismic pulse - thus the initial state, immediately before the pulse, may be assumed to be rest (or zero)). Once again our treatment is directed towards answering the question "how much can one do with the linearisation?" There will, of course, be systems where the influence of the parameters does not appear in the linear approximation (or is not recoverable therefrom); hence our methods, using the linear part, will provide no identification information.

It is clear from the proof of Lemma 4.5 that (A,C) initially observable + condition (4.26) gives initial observability for the augmented system. Thus we can envisage extension of previous work to look at parabolic systems and cases where α varies in space. One major constraint on this treatment, however, is the fact that the presence of unknown parameters in the highest order terms of an operator gives rise to "very unbounded" non-linearities (this is a reflection of the difficulties arising in perturbation theory when the perturbation is of the "same size" as the original operator). It is not yet clear how such problems should be handled within the present framework.

4.3 Adaptive control

In this section attention will be restricted to ordinary differential equations (i.e. finite dimensional systems). For such systems, problems of adaptive control often occur and have been studied by many authors. The adaptive controller is meant to compensate for the fact that, in the real world, perfect models are rarely available. Thus one tries some control and uses the observed response to that control in order to update one's knowledge of the controlled system and hence improve the control action. Of course there are many ways in which such adaptive procedures may be formulated. Here we consider a system of the form ($z \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $u \in \mathbb{R}^m$)

$$\begin{aligned} \dot{z} &= Az + Nz + Bu \\ y &= Cz \end{aligned} \qquad \ldots \qquad (4.29)$$

where the state z is regarded as containing unknown parameters (as in (4.20), (4.21)); these parameters embody our lack of confidence in our model of the real world - in as much as the real world is presumed to obey the same original model with possibly different parameter values. We shall assume that if the (augmented) state z is known at the beginning of an interval $[0,t_1]$ then the control action Bu is completely determined thereafter as a function of $z(0)$ and the output y - that is, we shall assume that if we knew the initial (unaugmented) state and the exact parameter values then we would be able to instantly calculate the feedback operator $F(z_0)$ which would provide the control action desired.

Thus the general scheme is as follows: consider positive time as having been split up into equal intervals $[jt_1,(j+1)t_1]$ where

$j = 0,1,2,\ldots$; $t_1 > 0$.

1. Produce (in a manner unspecified) an initial state guess z_0 ; set $j = 0$;

2. on an interval $[jt_1,(j+1)t_1]$, given an initial estimate $z_0(jt_1)$, apply the feedback $F(z_0(jt_1))$; this is modelled by

$$\begin{aligned} \dot{z} &= Az + Nz + F(z_0(jt_1))y \\ y &= Cz \ ; \end{aligned} \qquad \ldots \qquad (4.30)$$

3. now apply the parameter and state estimator of the preceding section (suitably adjusted for the known input $F(z_0(jt_1))y$) to formulate (and by successive approximation provide an algorithm for) the problem of reconstructing z over the interval $[jt_1,(j+1)t_1]$; let this fixed point formulation be denoted by $z = \Phi_j(z)$;

4. determine an initial guess over this interval by solving

   $\dot{z} = Az + Nz + F(z_0(jt_1))y$

   with $z(jt_1) = z_0(jt_1)$,

   (if, for instance, (4.30) results from a system which is linear when the parameters are known then this step is trivial);

5. then iterate Φ_j i-times to obtain a new approximation z_j^i on $[jt_1,(j+1)t_1]$; hence obtain a value $z_0((j+1)t_1)$ (as $z_j^i((j+1)t_1)$) ;

6. set $j: = j+1$ and return to 2 .

Steps 1. to 6. constitute our proposed adaptive control scheme. The next theorem indicates, rather crudely, the sorts of conditions one might impose to ensure that this scheme made sense i.e., ensure that our state and parameter estimates "converge" to the real world values. When this convergence occurs our control action will ensure that the real plant will behave as desired. Then we will say that an adaptive controller has been constructed. Thus we wish to ensure that on each interval $[jt_1,(j+1)t_1]$ iterating Φ_j brings us closer to the fixed point in $C(jt_1,(j+1)t_1;\mathbb{R}^n)$; and that moving from interval to interval in the fashion described does not upset convergence. The reason that only i iterations are allowed on each interval is the same as that for using adaptive control in the first place:- one is trying to adapt a controller, on-line, in real time, in response to observed plant behaviour. Thus it is not possible to iterate a large number of times on any particular interval.

Theorem 4.7

Recall Theorems 3.13 and 3.20, especially the latter. Assume that the linear part of the augmented system (4.30) is observable (i.e. (A,C) observable); Lemma 4.5 gives conditions for this. Assume also that $a,\beta_1,K,k_1,k_a,R,s,s',t_1 \in \mathbb{R}^+$, where $\frac{1}{s}+\frac{1}{s'}=1$, and that

a. $||e^{At} z|| \le g(t)||z|| \qquad t \ge 0 , z \in \mathbb{R}^n$

$g(\cdot) \in L^{s'}(0,t_1;\mathbb{R}^n)$

b. $N : C(0,t_1;\mathbb{R}^n) \to L^s(0,t_1;\mathbb{R}^n)$

continuous, $N(0) = 0$, and satisfies

$$||Nz - N\hat{z}||_{L^s(0,t_1;\mathbb{R}^n)} \leq k(||z||,||\hat{z}||)||z - \hat{z}||$$

where the norms on the right hand side are taken in $C(0,t_1;\mathbb{R}^n)$; $k : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, symmetric and such that $k(\theta_1,\theta_2) \to 0$ as $(\theta_1,\theta_2) \to (0,0)$.

c. $R > 0$ is such that

$$||C\int_0^{\cdot} e^{A(\cdot-s)}z(s)ds||_{\mathcal{Y}} \leq R||z||_{L^s(0,t_1;\mathbb{R}^n)}$$

where $\mathcal{Y}$ is the output space .

d. the feedback operator $F(v) : Y \to \mathbb{R}^n$ is such that

$$||F(v)y(\cdot)||_{L^s(0,t_1;\mathbb{R}^n)} \leq k_1||v||_{\mathbb{R}^n}||y(\cdot)||_{\mathcal{Y}}$$

e. taking $B_a = \{z \in C(0,t_1;\mathbb{R}^n) : ||z||_{C(0,t_1;\mathbb{R}^n)} \leq a\}$

we have that

$$(R||H_0^{-1}||_{L(\mathcal{Y},\mathbb{R}^n)}||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n))} + ||g(\cdot)||_{L^{s'}(0,t_1)})k_a \leq K < 1$$

where $k_a = \sup_{0 \leq \theta_1,\theta_2 \leq a} k(\theta_1,\theta_2)$

f. set $\beta_1 = \exp((k_a + 1)||e^{A\cdot}||_{C(0,t_1;\mathbb{R}^n)} t_1)$.

Then: the scheme described in steps 1. to 6., preceding the theorem, gives an adaptive control provided

$$||y(\cdot)|| \le \frac{a(1 - K)}{||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n))}||H_0^{-1}||_{L(Y,\mathbb{R}^n)} + k_1 a(R+||g||_{L^{s'}(0,t_1)})} \quad \ldots \quad (4.31)$$

and

$$\beta_1 K^i < 1 . \quad \ldots \quad (4.32)$$

Pf. Since the system (4.30) is autonomous we can reduce consideration of the scheme 1. to 6. to a series of problems defined over $[0,t_1]$, where the initial value for the next problem is obtained from the final value of the last. Note also that the difference between the maps Φ_j for different j lies solely in the term $F(\cdot)y$ which is fixed for a particular j.

Consider any Φ_j, the above remark allows us to conclude that it is a contraction on B_a viz.

$$\begin{aligned} ||\Phi_j z-\Phi_j \hat{z}||_{C(0,t_1;\mathbb{R}^n)} &\le ||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n))} \cdot ||H_0^{-1}||_{L(Y,\mathbb{R}^n)} \\ &\quad ||C\int_0^{\cdot} e^{A(\cdot-s)}(N(z(s)) - N(z(s)))ds||_Y + \\ &\quad ||g||_{L^{s'}(0,t_1)}||Nz - N\hat{z}||_{L^s(0,t_1;\mathbb{R}^n)} \\ &\le (R||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n))}||H_0^{-1}||_{L(Y,\mathbb{R}^n)} \\ &\quad + ||g||_{L^{s'}(0,t_1)})||Nz-N\hat{z}||_{L^s(0,t_1;\mathbb{R}^n)} \le K||z-\hat{z}||_{C(0,t_1;\mathbb{R}^n)} , \end{aligned}$$

on B_a, using b. and e.

Next we show that any such Φ_j maps B_a into itself. Choose $z \in B_a$; $v \in \mathbb{R}^n : ||v||_{\mathbb{R}^n} \leq a$ then we must consider

$$(\Phi_j z)(t) = e^{At} H_0^{-1}(y(\cdot) - C\int_0^{\cdot} e^{A(\cdot - s)}(Nz(s) + F(v)y(s))ds)$$

$$+ \int_0^t e^{A(t-s)}(Nz(s) + F(v)y(s))ds$$

$$||\Phi_j z||_{C(0,t_1;\mathbb{R}^n)} \leq ||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n))} ||H_0^{-1}||_{L(Y,\mathbb{R}^n)} (||y||_Y +$$

$$+ R||Nz||_{L^s(0,t_1)} + R||F(v)y||_{L^s(0,t_1)}) + ||g||_{L^{s'}} (||Nz||_{L^s(0,t_1)} +$$

$$+ ||F(v)y||_{L^s(0,t_1)}) \ .$$

Now using b. and d. this gives

$$\leq (||e^{A\cdot}||_{C(0,t_1;L(\mathbb{R}^n)} ||H_0^{-1}||_{L(Y,\mathbb{R}^n)} + Rk_1||z||_{C(0,t_1;\mathbb{R}^n)} +$$

$$+ k_1||g||_{L^{s'}(0,t_1)} ||z||_{C(0,t_1;\mathbb{R}^n)})||y||_Y + (R||e^{A\cdot}||_{C(0,t_1;\mathbb{R}^n)} ||H_0^{-1}||_{L(Y,\mathbb{R}^n)}$$

$$+ ||g||_{L^{s'}(0,t_1)}) \, k(||z||,0)||z||_{C(0,t_1;\mathbb{R}^n)} \ .$$

Thus for $z \in B_a$, $v : ||v||_{\mathbb{R}^n} \leq a$, we have, when $y(\cdot)$ satisfies (4.31),

$$\leq a(1-K) + Ka$$

$$\leq a \ .$$

Lastly we observe that if $z(0)$, $\hat{z}(0)$ are two initial conditions

$$z(t) - \hat{z}(t) = e^{At}(z(0)-\hat{z}(0)) + \int_0^t e^{A(t-s)}(N(z(s))-N(\hat{z}(s)))ds$$

gives that (by Gronwall's lemma)

$$||z - \hat{z}||_{C(0,t_1;\mathbb{R}^n)} \leq \beta_1 ||z(0)-\hat{z}(0)||_{\mathbb{R}^n} \quad \ldots \quad (4.33)$$

Now suppose we perform the sequence 1. to 6. with z_0 the very first guess at the state trajectory, lying in B_a. Let z^* denote the true trajectory. Then by the above i iterations of Φ_0 give a final state estimate $||z^i(t_1)-z^*(t_1)|| \leq K^i ||z_0-z^*||_{C(0,t_1;\mathbb{R}^n)}$.

Moving to the next interval we have that z_0 on the next interval is such that

$$\begin{aligned} ||z_0 - z^*|| &\leq \beta_1 ||z^i(t_1) - z^*(t_1)|| \\ &\leq \beta_1 K^i ||z_0 - z^*||_{C(0,t_1;\mathbb{R}^n)} \end{aligned}$$

(using (4.33) and step 4.)

Iterating i times on this interval we have that state error is reduced to $\beta_1 K^{2i} ||z_0-z^*||_{C(0,t_1;\mathbb{R}^n)}$. Thus the "reduction factor" is $\beta_1 K^i$ and condition (4.32) ensures that the whole adaptation process converges. ∎

The term $\beta_1 K^i$ expresses the balance between the amplification of errors in the initial state due to the natural dynamics of the system and the contraction properties of the observer iteration. The theorem tells us that, provided reality obeys a model of the form (4.30), then the adaptive control scheme (1. to 6.) will converge to the correct solution. This is typical of statements made about adaptive control algorithms by several other authors; though many of these works concern only systems which are linear in the state and linear in the parameters (but jointly bilinear). One could expect to expand upon the above result in such special cases.

The above result can also be extended to certain infinite dimensional systems. Indeed Theorem 4.7 is stated so as to be consistent with the (infinite dimensional) formulation of Chapter 3. The history and literature of adaptive control techniques, however, has been concerned with finite dimensional systems. The above result might also be extended to circumstances where the control action could not be immediately explicitly calculated (as above) but had to be determined by an iterative procedure. The control version of the present fixed point treatment could be used, for instance. One must then ensure that this joint procedure (iterating both for correct control and correct state) converges. It is not clear how best to formulate such an iteration. Lastly, in practice these results merely serve to ensure that certain procedures are reasonable; often one cannot expect to verify all the conditions before trying to apply the algorithm. The lack of a rigorous convergence analysis has not prevented the application of other adaptive control techniques.

CHAPTER V : Notes on constructive aspects

Summary

As stated in the introduction (Section 1.4) this chapter will be concerned with a largely formal account of some algorithmic possibilities arising from the treatment of Chapters 3 and 4. The intention is to provide some flavour of the "numerical analysis" which might arise from the preceding treatment, whilst avoiding over-burdening details and technical complexity. In cases where existence and (local) uniqueness of a fixed point is proven by a contraction argument a constructive procedure ("successive approximation") is automatically available. This is not so in cases where Schauder is used. One may then wish to know what would happen for successive approximation - or develop other iterations. Even in the contraction case one may wish to use other procedures in order to speed convergence.

5.1 Schauder : uniqueness

The uniqueness result of Kellogg:Smith-Stuart (Theorem 6, Appendix 4 : see Smith-Stuart [1]) is intended for application in cases where the existence of a fixed point has been proven by Schauder's theorem. In our case we shall consider the Φ of (3.28) and assume that the existence of a fixed point has been ensured by a theorem such as Theorem 3.23. For the result of Appendix 4 to apply one requires:

a. Φ is Frechet differentiable;

b. there is no fixed point of Φ on the boundary of B_a ;

c. the set $\{z \in B_a : 1$ is an eigenvalue of $d\Phi|_z\}$ has no accumulation points in B_a .

Assumption b. will be satisfied in the case of Theorem 3.23 for the norm bounds there give that $\Phi(B_a)$ in fact lies in the interior of B_a . Thus we are reduced to considering the Frechet derivative $d\Phi|_z$ for all points $z \in B_a$.

Results on the Frechet differentiability of Hammerstein operators (for which, see Martin [1]) give some indication of conditions on the semigroup and non-linearity which will ensure both continuous Frechet differentiability of Φ , and representation of the derivative in the desired form. Assuming some such result to hold we take ($v \in C(0,t_1;Z_1)$)

$$(d\Phi|_z(v))(t) = S(t)H_0^{-1}(- C\int_0^{\cdot} S(\cdot-s)(df|_z(v))(s)ds) + \int_0^t S(t-s)(df_z(v))(s)ds \quad \ldots \quad (5.1)$$

Now suppose we choose a point $z_1 \in B_a$ such that $d\Phi|_{z_1}$ has an eigenvalue of 1 . That is to say $v \in C(0,t_1;Z_1)$ such that

$$d\Phi|_{z_1}(v) = v .$$

From (5.1) we have, then,

$$v(t) = S(t)H_0^{-1}(-C\int_0^{\cdot} S(\cdot-s)(df|_{z_1}(v))(s)ds) + \int_0^t S(t-s)(df|_{z_1}(v))(s)ds \quad \ldots \quad (5.2)$$

Now consider the system

$$\dot{v} = Av + (df|_{z_1})(v) \quad \ldots \quad (5.3)$$

$$y = Cv \quad \ldots \quad (5.4)$$

Curtain-Pritchard [1] gives conditions under which the perturbed semigroup in (5.3) generates a mild evolution operator, and the (mild) solution of (5.3) can be represented as

$$v(t) = S(t)v_0 + \int_0^t S(t-s)(df|_{z_1}(v))(s)ds \quad .$$

Thus (5.2) corresponds to (using the same procedure as led to (3.28)) a reconstruction of the state for the linear system (5.3), (5.4), given an output which is zero. If the linearisation at z_1 ((5.3)) is observable (i.e. $(A + df|_{z_1}, C)$ is observable) then we may conclude that any such reconstructed state, v, must be identically zero. Conversely, if the linearisation $(A + df|_{z_1}, C)$ is unobservable then there will exist a non-trivial v satisfying (5.2). Hence, formally, we have

<u>Proposition 5.1</u>

A point $z_1 \in B_a$ is such that $d\Phi|_{z_1}$ has an eigenvalue of 1 iff the linearised system (5.3), (5.4) at z_1 is unobservable.

Pf. All results in this Chapter will be stated as summaries of (hopefully plausible) formal arguments. No rigorous proofs will be given.

Hence we may state

Proposition 5.2

Let Φ , f be as above and let Theorem 3.23 apply. *Then*: (3.28) has an *unique* solution in the ball B_a iff the set $\{z \quad B_a : (A + df|_z, C)$ is unobservable$\}$ does not have any limit points in B_a .

Pf. By Proposition 5.1 and Appendix 4, Theorem 6.

One might interpret points at which the linearised system is unobservable as points from which non-uniqueness may arise. That is to say, by looking only at the linear approximation around these points we cannot see in certain directions. Proposition 5.2 tells us that this local blindness does not prevent us from globally reconstructing the solution as long as there are not "too many" blind spots. Of course, there may be fixed points of Φ even when conditions a., b., c. do not hold - but then uniqueness, using these semilinear methods, cannot be guaranteed.

Many practical observation (resp. control) problems have curves, or surfaces, in the state trajectory space made up of points at which the linearisation is not observable (resp. controllable). It is arguable that the inability to ensure uniqueness in such problems is a good reason for not using this semi-linear approach. Non-unique reconstruction seems closely akin to the traditional notion of unobservability. This does not matter so much in the control case, where we are only interested in reaching some final state and not on how we get there. To study such questions one is forced to make more detailed analyses concerning the interaction between the system dynamics and "directions of blindness" that is, to consider approximations of higher order than linear. This is one aspect of non-linear geometric control theory, and is currently the subject of much research.

5.2 Successive approximation

Successive approximation, that is,

$$z_{i+1} = \Phi z_i$$

gives a particularly simple algorithm; if it converges then we obtain a fixed point of Φ . It is thus important to have available some general conditions (other than contraction) which will ensure, at least locally, convergence of this procedure. One answer is provided by

Proposition 5.3

Let Φ of (3.28) be a map from $C(0,t_1;Z_1)$ into itself. Let Φ, f satisfy the differentiability and representation assumptions of Section 5.1. Suppose that z_* is a fixed point of Φ and that $d\Phi|_{z^*}$ is a compact (linear map).

$$\text{Set } \sigma = \sup\{\lambda : (A + (\frac{df|_{z^*}}{\lambda}), C) \text{ is unobservable}\} \quad \ldots \quad (5.5)$$

If $\sigma < 1$ then z_* is a point of attraction for the iteration $z_{n+1} = \Phi(z_n)$.

Pf. By a slight modification of the discussion preceding Proposition 5.1 one has that the statement "$d\Phi|_{z_*}$ has $\lambda \neq 0$ as an eigenvalue" (as it is compact $d\Phi|_{z_*}$ only non-zero spectrum is point spectrum) is equivalent to the system

$$\dot{v} = Av + \frac{1}{\lambda}(df|_{z_*})(v) \quad \ldots \quad (5.6)$$

$$y = Cv$$

being unobservable. Thus the spectral radius requirement of the Ostrowski - (refined by) -Kitchen result of Appendix 4 becomes the condition stated as (5.5).

Of course $\sigma < 1$ ensures that 1 is not an eigenvalue of $d\Phi|_{z_*}$. Recall also that if Φ is a compact map and continuously Frechet differentiable then compactness of the Frechet derivative follows. Condition (5.5) concerns that "size" of perturbation needed to cause breakdown of observability - the bigger the perturbation required, the smaller σ will be. Intuitively Proposition 5.3 says "providing the linear part is dominant enough at the fixed point, Φ will be a local contraction there." This should be contrasted with Theorem 3.20 which does not use any differentiability assumptions.

Further, suppose thatthe hypotheses of Proposition 5.2 hold; the fact that for any starting point z_0 the sequence of successive iterates $z_1, z_2, z_3, \ldots$ lies in a compact subset of $C(0,t_1;Z_1)$, and therefore contains a strongly convergent subsequence does *not* allow us to conclude that the limit of this subsequence, is the desired fixed point (as would be the case if the sequence $z_1, z_2, \ldots$ itself were convergent). The existence of such convergent subsequences prompts a search for transformations F such that iteration of $F(\Phi)$ will converge to a fixed point of Φ. Consider such an $F(\Phi)$: in order to apply Theorem 7, Appendix 4, we must determine the spectrum of $d(F(\Phi))|_{z_*}$. To do this we should like to use the Gelfand calculus to obtain the spectrum of the Frechet derivative of the transformed Φ as the image under F of the spectrum of the Frechet derivative of Φ . Since the Gelfand calculus requires a complex Banach algebra we need extra hypotheses to ensure that we are justified in using the complexification technique. In addition an arbitrary F cannot be used - we must ensure that a fixed point of $F(\Phi)$ is a fixed point of Φ . It is not yet clear how best to perform these analyses.

5.3 The Newton method

For the map Φ of (3.2) we shall in this section consider, not the fixed point problem: find a z such that $\Phi(z) = z$, but the root finding problem: find a z such that $(I - \Phi)(z) = 0$. Such problems are traditionally solved by Newton's method. Following the formulation of Newton in Kantorovich-Akilov [1] and assuming differentiability and representation results as in Section 5.1, we have an iteration formally defined as

$$(I - d\Phi|_{z_n})^{-1}(\Phi - I)(z_n) + z_n = z_{n+1} \quad \ldots \quad (5.8)$$

If we define the defect

$$d_n = (\Phi - I)(z_n) \quad \ldots \quad (5.9)$$

and the update

$$v_n = z_{n+1} - z_n \quad \ldots \quad (5.10)$$

we get, by re-arranging (5.8),

$$(I - d\Phi|_{z_n})v_n = d_n \quad \ldots \quad (5.11)$$

the right-hand side of (5.11) becomes

$$d_n(t) = S(t)H_0^{-1}(y(\cdot)-\int_0^{\cdot} S(\cdot-s)f(z_n(s))ds) + \\ + \int_0^t S(t-s)f(z_n(s))ds - z_n(t) \quad \ldots \quad (5.12)$$

The left hand side of (5.11) gives

$$(I - d\Phi|_{z_n})v_n(t) = v_n(t) - \int_0^t S(t-s)(df|_{z_n}(v_n))(s)ds$$

$$S(t)H_0^{-1}(C\int_0^{\cdot} S(\cdot-s)(df|_{z_n}(v_n))(s)ds) \quad \ldots \quad (5.13)$$

Equating (5.12) and (5.13) gives

$$(v_n+z_n)(t) = z_{n+1}(t) = S(t)H_0^{-1}(y(\cdot)-C\int_0^{\cdot} S(\cdot-s)(f(z_n(s)) + $$
$$+ (df|_{z_n}(v_n))(s)ds) + \int_0^t S(t-s)(f(z_n(s)) + (df|_{z_n}(v_n))(s))ds \quad \ldots \quad (5.14)$$

Now using (5.10) we substitute $z_{n+1} - z_n$ for v_n and obtain

$$z_{n+1}(t) = S(t)H_0^{-1}(y(\cdot) - C\int_0^{\cdot} S(\cdot-s)(df|_{z_n}(z_{n+1}))(s)ds$$
$$- C\int_0^{\cdot} S(\cdot-s)(f - df|_{z_n})(z_n(s))ds) +$$
$$+ \int_0^t S(t-s)(df|_{z_n}(z_{n+1}))(s)ds +$$
$$+ \int_0^t S(t-s)(f - df|_{z_n})(z_n(s))ds \quad \ldots \quad (5.15)$$

As in Section 5.1 we may associate, at least formally, a sequence of linear problems with (5.15). For, consider the perturbed linear system

$$\begin{aligned} \dot{z}_{n+1} &= (A + df|_{z_n})(z_{n+1}) + (f - df|_{z_n})(z_n) \\ y &= Cz_{n+1} \end{aligned} \quad \ldots \quad (5.16)$$

Then proceeding as in Section 5.1 we may regard (5.15) as arising from the initial state reconstruction problem for (5.16). The iteration written as (5.16) can be regarded as repeated linearisation and will make sense if the pair $(A + df|_{z_n}, C)$ is observable. Such an iterative structure based on repeated linearisation is not uncommon elsewhere in non-linear estimation. The extended Kalman filter is a well known example. In (5.16), however, we have a known ("bias") correction term $(f - df|_{z_n})(z_n)$. Terms of this nature are often inserted in an ad hoc fashion.

It is possible to provide a convergence analysis, for this iteration, based on the theorems, concerning Newton's method, in Kantorovic-Akilov [1]. These theorems demand hypotheses on the first and second Frechet derivatives of Φ . The interpretation, in our case, looks somewhat inelegant and it might be better to attack the iteration (5.16) directly. These matters as already indicated are not pursued here and we shall conclude this Chapter with some remarks on extensions of the treatment in this section.

Iterated re-linearisation can be computationally onerous, but is often performed in off line design studies. Various authors have investigated the possibility of simplified versions. The simplest of these (again, see Kantorovich-Akilov, [1]) uses $d\Phi|_{z_0}$ in all the iterative steps (5.16) instead of $d\Phi|_{z_n}$. In case that $d\Phi|_{z_0} \equiv 0$ then this reduces to our contraction iteration. In any case, for computational purposes one is inevitably dealing with some approximate version. Certainly, when Φ is compact, $d\Phi|_{z_n}$ is compact linear and so can be approximated a finite dimensional operator.

Computational experience in other fields shows that the Newton algorithm may converge even when the conditions of Kantorovic-Akilov [1] fail. In particular one can, in some cases, by appropriately manipulating the iterative scheme, achieve convergence even to points lying on surfaces where the Frechet derivative is singular. In our case raises the interesting possibility of considering "intrinsically non-linear" observation problems. Typically such problems will possess curves, or surfaces, of points at which the linearised system will be unobservable; for instance, problems arising in satellite control show such phenomena. Further investigation of this point seems desirable since it offers a way of overcoming one of the main defects of the present treatment - the strength of the conditions on the linear part.

Appendix 1 : embeddings

For a region $\Omega \subset \mathbb{R}^n$, whose boundary $(\partial\Omega)$ is sufficiently smooth (explicitly : has the cone property) we define the Sobolev spaces $W^{m,p}(\Omega)$, where m,p are positive integers, as the space of real valued functions on Ω such that all derivations up to and including order m are L^p integrable. In this thesis $H^m(\Omega)$ denotes $W^{m,2}(\Omega)$. From Adams [1] we have that ...

1. bounded, $mp > n$; then the embeddings

   $W^{m,p}(\Omega) \to C^o(\Omega)$

   $W^{m,p}(\Omega) \to W^{o,q}(\Omega) \qquad 1 \leq q \leq \infty$

   are compact.

2. $mp > n$ implies that $W^{m,p}(\Omega)$ is a Banach algebra under pointwise products.

3. $mp \leq n$, bounded, j a positive integer

   $i: W^{j+m,p}(\Omega) \to W^{j,q}(\Omega)$ (where

   $0 < n - mp < n$ and $1 \leq q < \frac{np}{n-mp}$) is compact

   and

   $i: W^{j+m,p}(\Omega) \to W^{j,q}(\Omega)$ (where $n = mp$, $1 \leq q \leq \infty$)

   is compact.

Example: $W^{1,2}(\Omega) = H^1(\Omega) \to W^{o,q}(\Omega) = L^q(\Omega)$ is

compact if $n \geq 3$ and $1 \leq q < \frac{2n}{n-2}$.

Example: If $p > n$ $W^{1,p}(\Omega)$ is a Banach algebra

i.e. $H^1(\Omega)$ is a Banach algebra for $n = 1$.

If the Sobolev space is also a Banach algebra any polynomial (products being defined pointwise) will be well-defined. The norms of the embeddings noted above sometimes appear in calculations. In general optimal estimates for these constants are difficult to obtain; see Adams [1] and Lions-Magenes [1].

The following result will be found in Lions [3]. Take three Banach spaces B_0, B, B_1 with continuous embeddings $B_0 \to B \to B_1$; B_0, B_1 reflexive; and the embedding $B_0 \to B_1$ being compact. Define

$$W^{p_0,p_1}_{B_0,B_1} = \{v : v \in L^{p_0}(0,t_1;B_0),\ v' = \frac{dv}{dt} \in L^{p_1}(0,t_1;B_1)\}$$

where t_1 is finite and $1 < p_0, p_1 < \infty$. Equipped with the norm

$$||v||_{L^{p_0}(0,t_1;B_0)} + ||v'||_{L^{p_1}(0,t_1;B_1)}$$

$W^{p_0,p_1}_{B_0,B_1}$ is a Banach space. If $p_0 = p_1 = 2$ and B_0, B_1 are Hilbert then so is $W^{p_0,p_1}_{B_0,B_1}$ (when $B_1 = B_0^*$ we denote this space by $W_{B_0}(0,t_1)$) .

<u>Theorem</u> (Lions)

Under the above hypotheses the embedding of $W^{p_0,p_1}_{B_0,B_1}$ in $L^{p_0}(0,t_1;B)$ is compact.

Appendix 2 : analytic semigroups

Analytic semigroups can be regarded as an "operational" expression of the smoothing action created by parabolic partial differential equations. More precisely; let Z be a Banach space, $A:D(A) \to Z$ a closed, densely defined linear operator in Z. A is called *sectorial* if there are constants $\phi, M, a : 0 < \phi < \pi/2$, $M \geq 1$, $a \in \mathbb{R}$ such that the sector $S_{\phi,a} = \{\lambda \in C \mid \lambda \neq a, \phi < \arg|\lambda - a| \leq \pi\}$ is contained in $\rho(A)$, the resolvent set of A, and

$$||(\lambda - A)^{-1}|| \leq \frac{M}{|\lambda - a|} \quad \forall \lambda \in S_{\phi,a} \quad .$$

If A is sectorial then $k \geq 0$ such that $\operatorname{Re} \sigma(A+kI) > 0$. Let $A_1 = A + kI$. For $0 < \alpha < 1$ define

$$A_1^{-\alpha} = \frac{\sin \pi\alpha}{\pi} \int_0^\infty \lambda^{-\alpha}(\lambda + A_1)^{-1} d\lambda \; .$$

Then $A_1^{-\alpha}$ is bounded and injective. Let Z^α be the range of $A_1^{-\alpha}$, $Z^0 = Z$, $Z^1 = D(A)$; then we can take $A_1^\alpha : Z^\alpha \to Z$ to be the inverse of $A_1^{-\alpha}$, $A_1^0 = I_Z$, and $A_1^1 = A$. Z^α is dense in Z. Define the norm $||\cdot||_\alpha$ on Z^α by $||z||_\alpha = ||A_1^\alpha z||$ where $||\cdot||$ denotes the norm for Z. Z^α does not depend on the choice of k; different choices of k yield equivalent norms on Z^α. Z^α is a Banach space under $||\cdot||_\alpha$.

Example: Let Ω be an open bounded set in $\mathbb{R}^n$ whose boundary is of class C^{2m} (m an integer). Let $Z = L^2(\Omega)$,

$D(A) = H^{2m}(\Omega) \cap H_0^m(\Omega)$, $(Az)(x) = \sum_{|\alpha|\leq m} a_\alpha(x)(D^\alpha z(x))$
where the $a_\alpha : \bar{\Omega} \to \mathbb{R}$ are continuous mappings and $D^\alpha z$ is taken as a distributional derivative. Suppose that A is uniformly strongly elliptic on Ω , i.e.

$$\exists c_0 \in \mathbb{R}^+ : (-1)^m \sum_{|\alpha|=m} a_\alpha(x).\xi^\alpha \geq c_0|\xi|^{2m}$$

$\forall \xi = (\xi_\alpha)_{|\alpha|\leq m}$, $\xi\alpha \in \mathbb{R}$ and $\forall x \in \Omega$. Then A is sectorial. Indeed in this case $R(\lambda,A)$ is compact $\forall \lambda \in \rho(A)$.

If A is sectorial $-A$ generates an analytic semigroup. That is to say, a semigroup satisfying Definition 2.1 and, in addition, $t \to S(t)$ is real analytic on $]0,\infty[$ $\forall z \in Z$. Conversely we know that if $-A$ generates an analytic semigroup then A is sectorial. A simple expression of the smoothing property is

"Let A be sectorial, and $-A$ generate an analytic semigroup $S(t)$; m is any positive integer. Then $\forall t > 0$ $R(S(t)) \subset D(A^m)$."

Appendix 3 : pseudo-inverses

Further to the material contained in the first part of 2.2 we have (for $T \in L(X,Y)$ with closed range)

$$(T^*)^\dagger = (T^\dagger)^*$$

$$(T^\dagger)^\dagger = T$$

$$(T^*T)^\dagger = T\,(T^*)^\dagger$$

and so

$$T^\dagger = (T^*T)^\dagger T^* = T^*(TT^*)^\dagger \ .$$

It can also be shown that

$$N(T) = \{0\} \Rightarrow T^\dagger = (T^*T)^{-1}T^*$$

and

$$N(T^*) = \{0\} \Rightarrow T^\dagger = T^*(TT^*)^{-1} \ .$$

Suppose that $T = BC$ with B^*, C being surjective, then $T^\dagger = C^\dagger B^\dagger = C(CC^*)^{-1}(BB^*)^{-1}B^*$. The following 2 lemmas consider composition, and direct products, of maps.

Lemma 1

Let H_1, H_2, H_3 be Hilbert spaces; T_1 (resp. T_2) being bounded linear and with closed range from H_1 to H_2 (resp. H_2 to H_3) . Suppose that $R(T_2^*) \subset R(T_1)$ then $R(T_2 \circ T_1)$ is closed in H_3 .

Lemma 2

Let H_1, H_2, H_3 be Hilbert spaces; T being a bounded linear map from H_1 to H_2, S being a bounded linear map from H_1 to H_3, both T and S having closed range. Consider the map

$$T \times S : H_1 \to H_2 \times H_3 : u \to (Tu, Su) \quad ,$$

this has closed range iff the image of N(T) under S is closed in H_3 and the image of N(S) under T is closed in H_2.

Sometimes either the domain, or the range, has a particularly simple structure. Such structure can be used to advantage in the computation of the pseudo inverse.

Lemma 3

Let F be bounded linear $F : X \to X_F$ and onto (X, X_F both Hilbert), $T : X \to Y$ as before, $R(T^*) = R(F^*)$; then

$$T^\dagger = F^*(FT^*TF^*)^{-1}FT^* \, .$$

Lemma 4 (dual of 3)

Let E^* be bounded linear $Y \to Y_{E^*}$, and onto (Y, Y_{E^*} both Hilbert), $T : X \to Y$ as before and $R(T) = R(E)$ then

$$T^\dagger = T^*E(E^*TT^*E)^{-1}E^* \, .$$

Example: if T is represented by a matrix one might in Lemma 3 take H to consist of the linearly independent rows of T .

Appendix 4 : fixed point theorems

Many results have been developed for study of the fixed point problem. This appendix briefly summarises those results which are used in this thesis. Let Φ be a map from a Banach space X into itself. One of the first results was ...

Theorem 1 (Banach contraction : first form)

Let $\Phi : X \to X$ be such that

$$||\Phi x - \Phi\hat{x}|| \leq K||x - \hat{x}|| \quad \forall x_1, x_2 \in X$$

for some $K : 0 < K < 1$.

Then Φ has a fixed point in X .

The above theorem can be adapted so as to provide for "local" results (as in this thesis).

Theorem 2 (Banach contraction : second form)

$\Phi : X \to X$ as above. Let D be a closed subset of X and $||\Phi x - \Phi\hat{x}|| \leq K||x - \hat{x}||$, $\forall x_1, x_2 \in D$ for some $K \in]0,1[$. The iterative procedure ("successive approximation") $x_{i+1} = \Phi x_i$, $i = 0,1,2,...,$ converges to an unique solution in D of $\Phi x = x$ if the sphere

$$S = \{x \in X : ||x - x_1|| \leq \frac{K}{1-K} ||x_1 - x_0||$$

lies in D .

Generalising Brouwer's theorem in finite dimensions we have

<u>Theorem 3</u> (Schauder)

A continuous operator Φ which maps a closed convex subset, S, of X into a precompact subset of S, has a fixed point in S.

For operator splittings such as are discussed in Chapters 3 and 4 attempts to combine the properties of the Banach and Schauder theorems have been made.

<u>Theorem 4</u> (Nussbaum (see additional references).)

Let S be a closed bounded convex subset of the Banach space X. Suppose that Φ_1 and Φ_2 are continuous mappings from S into X such that

i) $(\Phi_1 + \Phi_2)S \subset S$

ii) $||\Phi_1 x - \Phi_1 \hat{x}|| \leq K||x - \hat{x}|| \qquad \forall x, \hat{x} \in S$

iii) $\overline{\Phi_2(S)}$ is compact

then $\Phi_1 + \Phi_2$ has a fixed point in S.

The results of Leray-Schauder [1] concern the application of Leray-Schauder degree to fixed point problems. The main result of that reference is

Theorem 5

Consider the equation

$$z - F(z,\mu) = 0 \quad . \quad . \quad . \qquad (A4.1)$$

under the following assumptions:

a. Z is a real Banach space with norm $||\cdot||$, $z \in Z$ and $F(\cdot,\cdot)$ takes values in Z .

b. The values of the parameter μ lie in an interval, M , on the real line ($|.|$ denotes absolute value).

c. $Z \times M$ denotes the product space with norm

$$||z - z'|| + |\mu - \mu'| \quad \text{for } z,z' \in Z \; ; \; \mu,\mu' \in M \; .$$

d. $F(z, \;)$ is defined on the closure $\overline{\Omega}$ of an open bounded set Ω in $Z \times M$.

e. $F(\cdot,\cdot)$ is compact on $\overline{\Omega}$ and uniformly continuous in μ .

f. $\partial\Omega$ does not contain any solution (z,μ) of (A4.1).

g. Or some $\mu_0 \in M$, (A.41) possesses a finite number of solutions all of which are known. Thus at μ_0 we may calculate the total Leray-Schauder index which we assume to be different from zero.

Then: we may conclude that there is a solution $\forall \mu \in M$ and, moreover, it is a solution which varies continuously with M .

We may wish to ensure uniqueness in cases where existence has been proven using Theorem 3.

Theorem 6 (Kellogg : Smith-Stuart)

Let Φ, S be as in Theorem 3. Suppose that

i) Φ is continuously Fréchet differentiable on S ;

ii) there is no fixed point of Φ on the boundary of S ;

iii) for each $z \in S$, 1 is not an eigenvalue of the Fréchet derivative at z of Φ (denoted $d\Phi|_z$) .

Then Φ has an unique fixed point in S .

The result also holds (for $\dim X > 1$) if iii) is replaced by

iii)' the set $\{z \in S :$ 1 is an eigenvalue of $d\Phi|_z\}$ has no points of accumulation in S .

Concerning the convergence of the sequence generated by successive approximation we have

Theorem 7 (Ostrowki : Kitchen [1] : Sermange)

Let f be a mapping whose domain and range are subsets of a Banach space X . Suppose that

i) $x^* \in X$ is a fixed point of f ;

ii) f is differentiable at x^* ;

iii) the spectral radius of the derivative of f at x^* is less than 1.

<u>Then</u>: there exists a neighbourhood N of x^* such that

$$\lim_{n\to\infty} f^n(x) = x^*$$

for each $x \in N$.

If such a neighbourhood N exists we shall say that x^* is a point of attraction for the iteration $x_{n+} = f(x_n)$.

List of notations

a	real constant
α	parameter vector
A	linear operator, possibly unbounded
$A(\cdot)$	family of same
B_a	ball radius a in some function space
C	operator giving output from state
$\mathcal{C}$	operator giving output trajectory from state trajectory (see (2.26))
$d\ldots\|_-$	Fréchet derivative of ... at $-$
$D(\cdot)$	domain of an operator
$f(\cdot)$	function : either of time or state z (usually appearing as non-homogeneous term in evolution equation)
H	Hilbert space
$H^m(\Omega)$	Sobolev space; see Appendix 1
H_0	initial state to output operator
$J(.)$	cost functional
K	real constant
L^p	p.th power Lebesgue integrable functions
$\mathcal{L}(X,Y)$	space of bounded linear operators from X to Y

$N(\cdot)$	the kernel of an operator
p, p_1, p_2	real numbers
Φ	non-linear operator
q	real number
r	real number
$R(\cdot)$	range of an operator
s	real number; dummy integration variable
$S(\cdot)$	semigroup
t	time variable; t_1 denotes specific instant
T	bounded linear operator between Banach spaces
τ	dummy integration variable
u	control input; values in U, trajectory lies in space $\mathcal{U}$
U	space of values taken by control
$\mathcal{U}$	space of input trajectories
V	Banach space
$W^{m,p}(\Omega)$	Sobolev space; see Appendix 1
$W(0,t_1)$	a space of functions defined on $[0,t_1]$; see Definition 2.19
$W_z(0,t_1)$	a variation on the above

X	Banach space
Y	Banach space
$\mathcal{Y}$	space of output trajectories
z	is used to denote the system state
z_0	the initial state i.e $z(0)$
Z	the Banach space in which the system state lies
$*$	is used to denote adjoint space a d also operator
$\perp$	used to denote (in Hilbert space) "subspace orthogonal to ..."
$\rightarrow$	map or "has limit"; apparent from context
$\dagger$	denotes generalized inverse; see Definition 2.23
Δ	Laplacian
$\nabla.$	grad
∇	div

In the case of function valued function spaces e.g. $L^2(0,t_1;Z)$ $0,t_1$ denotes time interval $[0,t_1]$ and Z the Banach space in which these functions take their values.

References

Adams, R.A. : [1], Sobolev spaces, New York, Academic Press, 1975

Ball, J.M. : [1], Remarks on blow-up and nonexistence theorems for non-linear evolution equations, Quart. Jnl. Math., 28, p.473-486, 1977.

Ben-Israel, A.; Greville, T.N.E. : [1], Generalized inverses : theory and applications, New York, Wiley, 1974.

Beutler, F.J. : [1], The operator theory of the pseudo-inverse, (I,II), Jnl. Math. Anal. Appl., 10, p.451-470, p.471-493, 1965

Carmichael, N.; Pritchard, A.J.; Quinn, M.D. : [1], State and parameter estimation for non-linear systems, Control Theory Centre report No. 89, to appear in Appl. Math. and Opt.

Carmichael, N.; Pritchard, A.J.; Quinn, M.D. : [2], Non-linear control and estimation : an approach via fixed point theorems and degree theory, Control Theory Centre report No. 98, to appear in Proc. 1981 Graz symposium.

Carmichael, N.; Quinn, M.D. : [1], Semi-linear control theory - a survey, to appear.

Céa, J. : [1], Optimisation : théorie et algorithmes, Paris, Dunod, 1971

Chavent, G. : [1], About the stability of the optimal control solution of inverse problems, ZRIA note, 1979

Curtain, R.; Pritchard, A.J. : [1], Infinite dimensional linear systems theory, Berlin, Springer-Verlag, 1978

Dolecki, S.; Russell, D.L. : [1], A general theory of observation and control, S.I.A.M. Jnl. Control and Opt., 15, p.185-220, 1977

Falb, P.L.; Jong, L.d. : [1], Some successive approximation methods in control and oscillation theory, New York, Academic, 1969

Fujita, H. : [1], On some non-existence and non-uniqueness theorems for non-linear parabolic equations, Proc. Symp. on Pure Math., A.M.S., 18, p.105-113, 1970

Haraux, A. : [1], Non linear evolution equations - global behaviour of solutions, Berlin, Springer-Verlag, 1981

Henry, D. : [1], Geometric theory of semilinear parabolic equations, Berlin, Springer-Verlag, 1981

Hille, E. ; Phillips, R.S. : [1], Functional analysis and semigroups, Rev. Ed. A.M.S. Coll. Pub., 1957

Ichikawa, A.; Pritchard, A.J. : [1], Existence, uniqueness and stability of non-linear evolution equations, J. Math. Anal. Appl. 68, p.454-476, 1979

Jazwinski, A. : [1], Stochastic processes and filtering theory, New York, Academic, 1970

John, F. : [1], Blow-up of solutions of non-linear wave equations in three space dimensions, Man. Math., 28, p.235-268, 1979

Kantorovic, L.W.; Akilov, G.R. : [1], Functional analysis in normed spaces, Moscow, Nauka, 1959

Kitchen, J. : [1], Concerning the convergence of iterates to fixed points, Stud. Math., 27, p.247-249, 1966

Leray, J.; Schauder, J. : [1], Topologie et equations fonctionelles, Ann. Sc. E.N.S., 51, p.45-78, 1934

Lions, J.L. : [1], Equations differentielles, operationelles; et problemes aux limites, Berlin, Springer-Verlag, 1961

Lions, J.L. : [2], Contrôle optimale de systèmes gouvernés par des equations aux derivées partielles, Paris, Dunod, 1968

Lions, J.L. : [3], Sur quelques méthodes de résolution des problèmes aux limites non linéaires, Paris, Dunod, 1969

Lions, J.L. : [4], Function spaces and optimal control of distributed parameter systems, Rio de Janeiro Lecture Notes, 1980

Lions, J.L.; Magenes, E. : [1], Problèmes aux limites non homogènes et applications, Vols. 1,2,3, Paris, Dunod, 1968

Lions, J.L.; Stampacchia, G. : [1], Variational inequalities, C.P.A.M., 20, p.493-519, 1967

Magnusson, K. : [1], Controllability for non-linear systems, Ph.D. thesis, University of Warwick, 1982

Martin, R.H. : [1], Non linear operators and differential equations in Banach spaces, New York, Wiley, 1976

Nashed, M.Z. : [1], Generalized inverses, normal solvability and iteration for singular operator equations in "Non linear functional analysis and applications", Ed. L.B. Rall, New York, Academic, 1971

Protter,M.H.; Weinberger, H. : [1], Maximum principles in differential equations, New Jersey, Prentice-Hall, 1967

Reed, M. : [1], Abstract non-linear wave equations, Berlin, Springer-Verlag, 1976

Ribière, G. : [1], Régularisation d'opérateurs, R.I.R.O., 1, p.57-79, 1967

Segal, I. : [1], Non-linear semi-groups, Ann. Math., 78, p.339-364, 1963

Smith, H.L.; Stuart, C.A. : [1], A uniqueness theorem for fixed points, Proc. A.M.S., 79, p.237-240, 1980

Tikhonov, A.N. : [1], The regularisation of incorrectly posed problems, Dokl. Akad. Nauk. S.S.R., 153, p.42-52, 1963

Additional references.

Burgers, J.: A mathematical model illustrating the theory of turbulence, Adv. Appl. Mech., Vol. 1, 1948

Dias, J.P.: On the existence of a strong solution for a nonlinear evolutions system, Journal of Nonlinear analysis.

Dieudonne, J.: Treatise on analysis, Vol. II, Academic Press. New York, 1976

Kato, T.; Fujita, F., On the Navier Stokes initial value problem, Arch. Rat. Mech. Anal., 16, p. 269-315, 1964

Nussbaum, R.P.: The fixed point index and asymptotic fixed point theorems for K-set contractions, B.A.M.S., 75, p. 490-495, 1969

Weissler, F.B.: Local existence and non-existence for semilinear parabolic equations in L^p, Indiana Univ. Math. Jnl., 29, p. 79-102, 1980