Analog Computability with Differential Equations

Analog Computability with Differential Equations

By Diogo Poças, B.Sc., M.Sc.

A Thesis Submitted to the Department of Mathematics and Statistics and the School of Graduate Studies of McMaster University in Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy

Doctor of Philosophy in Mathematics (2017)

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada

TITLE:	Analog Computability with Differential Equations
AUTHOR:	Diogo Poças, B.Sc. (Universidade de Lisboa, Portugal), M.Sc. (Universidade de Lisboa, Portugal)
SUPERVISOR:	Jeffery I. Zucker
NUMBER OF PAGES:	x, 130

Abstract

In this dissertation we study a pioneering model of analog computation called General Purpose Analog Computer (GPAC), introduced by Shannon in 1941. The GPAC is capable of manipulating real-valued data streams. Its power is characterized by the class of differentially algebraic functions, which includes the solutions of initial value problems for ordinary differential equations.

We address two limitations of this model. The first is its fundamental inability to reason about functions of more than one independent variable (the 'time' variable). In particular, the Shannon GPAC cannot be used to specify solutions of partial differential equations. The second concerns the notion of approximability, a desirable property in computation over continuous spaces that is however absent in the GPAC.

To overcome these limitations, we extend the class of data types by taking channels carrying information on a general complete metric space X; for example the class of continuous functions of one real variable. We consider the original modules in Shannon's construction (constants, adders, multipliers, integrators) and add two new modules: a differential module which computes spatial derivatives, $u(t) \mapsto \partial_x u(t)$; and a continuous limit module which computes limits, $u(t) \mapsto \lim_{t\to\infty} u(t)$.

We then build networks using X-stream channels and the abovementioned modules. This leads us to a framework in which the specifications of such analog systems are given by fixed points of certain operators on continuous data streams, as considered by Tucker and Zucker. We study the properties of these analog systems and their associated operators. We present a characterization which generalizes Shannon's results. We show that some non-differentially algebraic functions such as the gamma function are generable by our model. Finally, we attempt to relate our model of computation to the notion of tracking computability as studied by Tucker and Zucker.

Acknowledgements

I would like to express my gratitude to my advisor, Jeff Zucker, for his invaluable dedication and guidance during the elaboration of this thesis. I also want to thank him for providing a space that allowed me to grow as a young researcher, and for the great patience and encouragement offered in the countless times we went through the several stages of this work.

I am also lucky for having Profs. Dmitry Pelinovsky and Bartek Protas on my advisory committee. They have provided me with insightful advices and feedback. I truly feel that our discussions were very helpful in improving the quality of this thesis.

I would like to thank my colleagues and all people in the Department of Mathematics and Statistics, for their support and assistance provided when needed, and for the opportunity to present some of my results at the AIMS Lab.

Thanks also to my former supervisor at IST, José Félix Costa, for his advice, support and friendship. I am also grateful to Daniel Graça and Amaury Pouly for their useful suggestions and appreciations of a portion of this work.

Moving to Canada was certainly a challenging experience at first, and I want to express my gratitude to all the wonderful people that helped Hamilton feel like a second home. I am lucky to have met Rita, Ana and Pedro, for immediately making sure I received a warm welcome and for our quasiweekly lunches.

I would like to thank Amay, Bingying and Chengwei for the adventures together. I am also grateful to my cheerful housemates Yusuke, Nadeem, Sam, Wesley, Peter, Tyler and Craig for their friendship and great memories; whether we were going to the Art Crawl or the Portuguese bar, playing boardgames or throwing barbecues, we were always having the best of times.

I would like to thank MACSA for their activities and for allowing me to grow spiritually in parallel with my academic life. I am especially glad for all the coffee breaks with Matt Chan.

I am thankful to my oldest friends in Portugal - in Linda-a-Velha, at Batalha, at IST, and the Cenáculo group, for their true friendship and enthusiasm. I hope our bonds stay steady for years to come.

Finally, I wish to give my very special thanks to my parents and my family, for their presence, support, interest and unconditional love.

Contents

List of Figures

Chapter 1

Introduction

1.1 History of analog computation

Analog computation, as conceived by Kelvin [TT80], Bush [Bus31], and Hartree [Har50], is a form of experimental computation with physical systems called analog devices or analog computers. Historically, data are represented by measurable physical quantities, including lengths, shaft rotation, voltage, current, resistance, etc., and the analog devices that process these representations are made from mechanical, electromechanical or electronic components [Sma93, Hol96, Joh96].

A general purpose analog computer (GPAC) was introduced by Shannon [Sha41] as a model of Bush's Differential Analyzer [Bus31]. Shannon discovered that a function can be generated by a GPAC if, and only if, it is differentially algebraic, but his proof was incomplete. A basic study was made by Pour-El [PE74] who gave some good characterizations of the class of analog computable functions, focusing on the classical analog systems built from constants, adders, multipliers and integrators. This yielded a stronger model and a new proof of the Shannon's equivalence (and some new gaps, corrected by Lipshitz, Rubel [LR87], Graça and Costa [GC03]). Using this characterization in terms of algebraic differential equations, Pour-El showed that not all computable functions on the reals (in the sense of computable analysis) can be obtained with these analog networks. The typical counterexample is the gamma function

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

which is not differentiable algebraic and so cannot be generated by a GPAC, as noted by Shannon himself. However, one could expect that, in a 'sensible' model of computability on continuous data, this function would be computable.

Indeed, the gamma function is *effectively computable* in the sense of computable analysis, a branch of constructive mathematics studied by Grzegorczyk [Grz55, Grz57], Lacombe [Lac55a, Lac55b, Lac55c], Pour-El, Richards [PER79], Weihrauch [Wei00], Tucker, Zucker [TZ07], among others. These authors have in one way or another tried to answer what is perhaps the fundamental question for analog computation: which functions are computable? For the case of digital computation, all empirical evidence corroborates the celebrated Church-Turing Thesis, showing that various models (such as Turing machines, λ-calculus and recursive functions) are equivalent. This picture is not so clear for analog computability on continuous spaces, despite the abundance of progress made and (partial) equivalence results [BCGH06, Ko91, SHT99, Wei00].

Returning to the Shannon GPAC and the (non-)computability of the gamma function, some

authors have attempted to include *approximability* in the model (which is an important ingredient in many models of real computation such as computable analysis); in particular, Graça [Gra04] redefined the notion of GPAC-computability in order to show that the gamma function can indeed be considered as GPAC-computable.

There is, however, another limitation with the Shannon GPAC which appears to have been overlooked by Shannon, Pour-El and others. It lies in the fact that the Shannon GPAC can fundamentally reason only about real-valued functions of one independent variable t. Ironically, it was stated in [Sha41] and [PE74] that the generalization to more than one independent variable only requires an obvious modification, but this is by no means the case. In fact, it is hard to conceive a realistic physical interpretation for a formalism involving two (or more) independent "time" variables.

To address this problem, and inspired by the assumption that the brain is a type of analog computer, Rubel defined the Extended Analog Computer (EAC) in [Rub93]. Rubel stressed that his model is a *conceptual* computer - the extent to which it can be realized by actual physical, chemical, or biological devices or systems would remain to be investigated. An implementation of the EAC (or at least, of some of its components) has been achieved with the work of Mills [Mil08].

The theory of analog computing has also been developed by Moore with some very general mathematical models [Moo96]. These models, using schemes rather like Kleene's [Kle55], but with primitive recursion replaced by integration and others added, define a hierarchy of functions on the reals, which contains the GPAC generable functions at its lowest level, and non-computable functions (in the sense of computable analysis) at higher levels. Graça and Costa [GC03] have presented an improved model of the GPAC, and shown this to be equivalent to the lowest level subclass of Moore's functions.

The contributions of Campagnolo [CMC02] and Mycka [MC04] have also presented some fine results concerning analog complexity classes. Finally, Pouly [Pou15] studied in his PhD thesis the GPAC (among other models of computation) from the point of view of complexity classes, and he, Bournez and Graça [BGP16] have defined a type of multidimensional GPAC.

1.2 Analog networks and evolution problems

The main objects of our study are *analog networks* or *analog systems* [TZ07, TZ11, JZ13, TZ14], whose main components are described as follows:

$$\textit{Analog network} = \textit{data} + \textit{time} + \textit{channels} + \textit{modules}.$$

We model *data* as elements of a complete metric vector space $\mathcal{X}$, such as a Banach or Fréchet space. We use a continuous model of *time* as an interval of the real numbers, either bounded ($\mathbb{T} = [0, T]$, where T denotes the final time) or unbounded ($\mathbb{T} = [0, +\infty)$). Each *channel* carries, for example, a continuously differentiable stream, represented as a function $u : \mathbb{T} \to \mathcal{X}$ (this space is denoted by $C^1(\mathbb{T}, \mathcal{X})$). Each *module* M has zero, one or more input channels, and must have a single output channel; thus it can be specified by a (possibly partially defined) *stream function*

$$F_M : \mathcal{X}^k \times C^1(\mathbb{T}, \mathcal{X})^\ell \rightharpoonup C^1(\mathbb{T}, \mathcal{X}); \quad k, \ell \in \mathbb{N}.$$

Our goal is to develop and extend the existing concepts into spaces of functions of several variables. As mentioned above, the GPAC can only reason about functions of one variable, and thus it is limited to initial value problems of ordinary differential equations. We can think of the GPAC as a particular example of analog network in which data correspond to the space of real numbers, $\mathcal{X} = \mathbb{R}$. Its limitations can be removed by assuming a more general data space $\mathcal{X}$. For example, we can think of $\mathcal{X}$ as the space of continuous real-valued functions on a bounded domain $\Omega \subset \mathbb{R}^n$, that

is, $\mathcal{X} = C(\Omega, \mathbb{R})$. In this way, our channels will now carry $\mathcal{X}$-valued streams of data $u : [0, T] \to \mathcal{X}$, which correspond to functions of $n + 1$ real variables, under the uncurrying

$$[0, T] \to (\Omega \to \mathbb{R}) \quad \simeq \quad [0, T] \times \Omega \to \mathbb{R}.$$

Our approach is, to some extent, motivated by the theory of partial differential equations, in which some fundamental problems (such as the heat equation, wave equation and Schrödinger equation) can be expressed as time evolution problems in a function space. An important disclaimer is that our approach is not concerned with applying the GPAC (or other computability models in general) in order to solve problems in PDEs, but the other way around; we intend to apply results and techniques from the theory of PDEs into models of computation such as the GPAC.

The idea of using a general metric space $\mathcal{X}$ for data is not new and, in fact, it is mentioned in the original papers [TZ07, TZ11, JZ13, TZ14] as well as in James' thesis [Jam12]. Their techniques rely on causality and contractivity properties of the network operators, which are sufficient to prove existence, uniqueness, continuity and computability of fixed points. The results obtained are indeed desirable (for any worthwhile model of computation); however, it turns out that their assumptions (namely, contractivity of the network operators) do not hold in the more general scenario that we wish to consider.

In informal terms, partial differential equations are just much more complicated than ordinary differential equations. Even for the most simple case of the transport equation $\partial_t u = \partial_x u$, the corresponding network operator is not contracting (in the usual topology); it even fails to be continuous! Moreover, the typical construction based on an iteration scheme to produce fixed points does not work in general.

As mentioned above, our study is directed towards functions of several variables and a multidimensional GPAC. As a technical remark, it turns out that some of the properties of the underlying spaces are no longer present, such as the existence of a norm. Instead of considering Banach spaces, as was done before in the literature, we look at Fréchet spaces (which come equipped with a family of pseudonorms). Actually, the notion of Fréchet space was present implicitly in the original papers of Tucker and Zucker. Our results provide evidence that this more general type of spaces is indeed the correct framework for our study. Even though working with Fréchet spaces is somewhat more technically demanding, some of the results in Banach spaces can be adapted into this case.

1.3 Outline of the thesis

We now provide a short summary of each chapter, with emphasis on the main original results.

In Chapter 2, we study linear evolution problems in the theory of analog networks. We explain how to view the solution of a linear evolution problem as the fixed point of an analog network. After dealing with the easy case in which data lie in a Banach space and the linear operator is bounded, we move into the more interesting case in which data lie in a Fréchet space. This turns out to be a necessary choice for the problem at hand. We then pursue two approaches, both motivated by the attempt to produce fixed points via iterating sequences. In the first approach, we establish parallels with the Cauchy-Kowalevski theory, which relies on analyticity assumptions; the main results are Theorems 7 and 10 which establish convergence of iterates to a fixed point. In the second approach, we apply the Fourier transform and study the corresponding networks in Fourier space; the main results are Theorems 12 and 13 which show existence, uniqueness and convergence of the fixed points.

In Chapter 3, we introduce the $\mathcal{X}$-GPAC, a generalization of the classical GPAC for functions of more than one variable. We take the original modules in Shannon's construction and add a *differential module* which produces spatial (partial) derivatives. We formalize the notion of *generable*

functions, originally present in Shannon's construction, and adapt it to our setting. As a technical point in defining the semantics of our networks, we use a *closedness condition* rather than continuity, and thus a weaker form of well-posedness (which we call quasi-well-posedness). We also consider a more abstract, multityped GPAC and present various modular operations such as module derivation and channel contraction. Our main result is Theorem 17, in which we characterize the class of $\mathcal{X}$-GPAC-generable functions, by defining (and obtaining a correspondence with) the class of solutions to *partial differential algebraic systems of equations*.

In Chapter 4, we attempt to incorporate *approximability* into the GPAC model of computation. This is achieved by introducing yet another module that produces effective limits. The main motivation of this chapter is to show how some classically non-generable functions, such as the gamma function and the Riemann zeta function, can be captured with the 'limit $\mathcal{X}$-GPAC', or LGPAC. This is achieved in Theorems 20 and 21, which constitute the main results of the chapter.

In Chapter 5, we connect the model of computation described in this thesis with other well-known models of computable analysis. In particular, we study the notion of *tracking computability* presented in [TZ04]. This construction starts from an enumeration of a countable dense subset of our underlying set, and defines computable elements as those given by effective Cauchy sequences. In this way, the notion of computability in continuous spaces can be translated into computability on the natural numbers by considering *tracking functions*. We show that all the relevant modules studied in this thesis induce computable tracking functions. Our main results are Theorems 22 and 23 which prove tracking computability of the induced operator and the semantics operator of an LGPAC.

Chapter 2

Linear Evolution Problems

In this chapter we show how to frame linear evolution problems (i.e. time evolution problems with a linear operator) in the theory of analog networks. The main goal is to present some concepts, notation and results that will appear consistently in the following chapters. The reason for choosing linear evolution problems is because they have a simple formulation and a well-known solution given by an exponential operator, so that the interesting concepts related to analog networks will be clearly identifiable.

We shall begin by introducing the linear evolution problem and convert it into the framework of analog networks. We introduce the concept of Fréchet spaces as a necessary tool for studying this problem in infinite dimensions. We reframe our problem as a fixed point problem and then follow two different approaches towards solving it. First, we use a Cauchy-Kowalevski approach, in which contraction inequalities are sought and then used in showing local and global convergence results. Then, we use a Fourier Transform approach, where contraction inequalities also play a role, but a different technique involving Fourier Transform approximations is utilised.

We now briefly comment on the original content of this chapter. We present a generalization to contraction inequalities (which is often used in literature for norm or metric functions; see Definition 2.1.4 and Theorem 2) in terms of *pseudonorms* in our Lemma 2.4.1. Even though this generalization seems immediate, we claim that (to the best of our knowledge) this is an original idea. This technique allows us to obtain original Theorems 7 and 10, which are to some extent covered by the Cauchy-Kowalevski Theorem; however, our results offer a constructive approach to obtain fixed points via iteration, which is not provided by that theorem. In the Fourier Transform approach, our main original results are Theorems 12 and 13; these are constructive versions of results covered by standard Fourier analysis, but translated to our framework of Fréchet spaces. Hence, they can be regarded as illustrating the power of contraction inequalities techniques when applied to frequency spaces.

2.1 The linear evolution problem

The setting for this chapter is as follows. We consider a complete metric vector space, denoted by $(\mathcal{X}, d)$ or simply by $\mathcal{X}$, and unbounded time interval $\mathbb{T} = [0, \infty)$. We also consider a linear operator L of type $\mathcal{X} \rightharpoonup \mathcal{X}$ and denote the domain of L by $D(L)$. The symbol $\rightharpoonup$ is used for partial-valued functions, meaning that $D(L) \subseteq \mathcal{X}$; when the function is known to be total, that is, $D(L) = \mathcal{X}$, we may use $\to$ instead. By a *stream function* we simply mean an element $u \in C(\mathbb{T}, \mathcal{X})$.

We recall that a *Banach space* is a complete metric vector space $(\mathcal{X}, d)$ in which the metric is

induced by some norm $\|\cdot\|_{\mathcal{X}}$.

Definition 2.1.1 (**Bounded / Unbounded operator**). Let $\mathcal{X}$, $\mathcal{Y}$ be Banach spaces and let $L : \mathcal{X} \rightharpoonup \mathcal{Y}$ be a linear operator. We say that L is *bounded* if there exists $C \in \mathbb{R}$ such that, for all $x \in D(L)$, we have $\|Lx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$; otherwise, we say that L is *unbounded*. When L is a bounded operator, we define the *norm of* L as the least nonnegative $C \in \mathbb{R}$ such that $\|Lx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$ for all $x \in D(L)$; this can be given as

$$\|L\| = \sup_{\substack{x \in \mathcal{X} \\ x \neq 0}} \frac{\|Lx\|_{\mathcal{Y}}}{\|x\|_{\mathcal{X}}} = \sup_{\substack{x \in \mathcal{X} \\ \|x\|=1}} \|Lx\|_{\mathcal{Y}}.$$

Example 2.1.2. Let $\mathcal{X} = C([0,1])$ be the space of continuous functions in the compact interval $[0,1]$ with the supremum norm, and let $L = \partial_x$ be the first derivative operator, which is a linear operator with $D(L) = C^1([0,1])$. Consider the sequence $a_n(x) = \frac{1}{n}\sin(n^2 x)$, which converges uniformly (that is, in the supremum norm) to 0, $\|a_n\| = \frac{1}{n} \to 0$. Moreover, each $a_n \in C^1([0,1])$, and $L(a_n)(x) = a_n'(x) = n\cos(n^2 x)$. Note that $\|a_n'\| = n$, meaning that the suprema of a_n' grow without bounds. Thus L is an example of an unbounded operator.

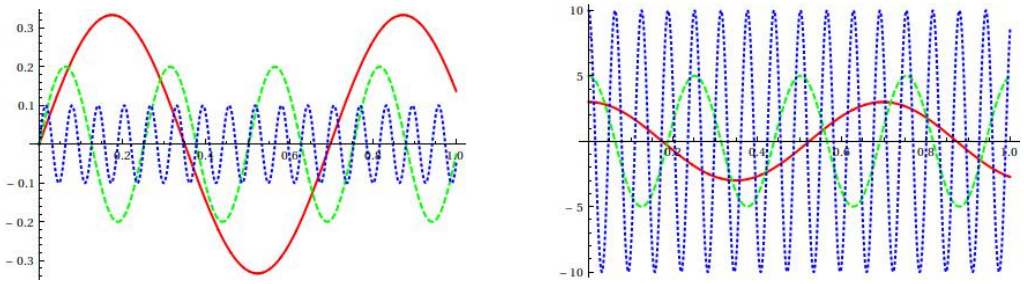


Figure 2.1: Plot of $a_n(x)$ (left) and $a_n'(x)$ (right) for $n = 3$ (red, bold), $n = 5$ (green, dashed), $n = 10$ (blue, dotted).

Even though our definition allows bounded operators to be partially defined, an important result shows that we may extend bounded operators to the whole space $\mathcal{X}$.

Theorem 1 (**Hahn-Banach Theorem, [RS80]**). *Let $L : \mathcal{X} \rightharpoonup \mathcal{Y}$ be a bounded operator. Then there exists a linear operator $\tilde{L} : \mathcal{X} \to \mathcal{Y}$ such that*

- *$\tilde{L}$ is total, that is, $D(\tilde{L}) = \mathcal{X}$;*
- *$\tilde{L}$ is an extension of L, that is, for all $x \in D(L)$ we have that $\tilde{L}x = Lx$;*
- *$\tilde{L}$ is bounded, and moreover, $\|\tilde{L}\| = \|L\|$.*

Thus, whenever considering bounded operators, we shall assume without loss of generality that they are total. No similar result holds for unbounded operators, and in fact it turns out that most unbounded operators of interest are partially defined, such as the derivative operator.

The main problem we wish to study in this chapter can be formulated as follows.

Definition 2.1.3 (**Linear evolution problem, [Had52]**). Let $\mathcal{X}$ be a complete metric vector space and $L : \mathcal{X} \rightharpoonup \mathcal{X}$ a linear operator (not necessarily bounded). For a given initial condition

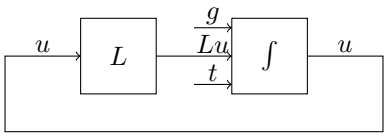


Figure 2.2: Analog network (with two modules and three channels) for the linear evolution problem.

$g \in \mathcal{X}$, the *linear evolution problem* (also called time evolution problem) is given by the system

$$\begin{cases} \frac{du}{dt} = Lu, & t \in \mathbb{T}; \\ u(0) = g. \end{cases} \tag{2.1}$$

A (strong) solution of the time evolution problem is a stream function $u \in C(\mathbb{T}, \mathcal{X})$ such that $u \in C(\mathbb{T}, D(L)) \cap C^1(\mathbb{T}, \mathcal{X})$ and u satisfies (2.1).

Sometimes we may be interested in finding a solution only in a bounded interval $[0, T]$, where $T \in \mathbb{T}$. When that happens we may refer to the desired $u \in C([0, T], \mathcal{X})$ as a *finite time solution*.

In this thesis we are interested in the formulation of time evolution problems (either with linear or nonlinear evolution operators) in the framework of analog networks. To construct an analog system that reasons about the linear evolution problem, we can simply integrate the differential equation (2.1) to obtain

$$u(t) = g + \int_0^t Lu(s)ds =: \Phi_L(g, u)(t), \tag{2.2}$$

where we use the right hand side to define an operator $\Phi_L : \mathcal{X} \times C(\mathbb{T}, \mathcal{X}) \rightharpoonup C(\mathbb{T}, \mathcal{X})$, which can be computed using an analog network with two modules. Introducing a feedback to implement the equality, we obtain the analog system of Figure 2.2.

For a given $g \in \mathcal{X}$, we can consider the *section of* Φ_L

$$\begin{array}{rrl} \Phi_{L,g} : & C(\mathbb{T}, \mathcal{X}) & \rightharpoonup C(\mathbb{T}, \mathcal{X}) \\ & (u, t) & \mapsto g + \int_0^t Lu(s)ds. \end{array} \tag{2.3}$$

We can then observe the equivalence between the notions of (a) solutions to the linear evolution problem of Definition 2.1.3; (b) specifications of the analog system of Figure 2.2; and (c) fixed points of the operator $\Phi_{L,g}$ defined by (2.3). Henceforth we will focus on the last notion. Our goal is to provide sufficient conditions that ensure existence and uniqueness of fixed points, as well as the existence of a constructive method to obtain fixed points when they exist.

A relevant result, which can be seen as a starting point for the discussion in this chapter, is the Banach fixed point theorem (also called the contraction mapping principle), which we state and prove.

Definition 2.1.4 (Contraction mapping). Let $(\mathcal{X}, d)$ be a metric space and $F : \mathcal{X} \to \mathcal{X}$. We say that F is a *d-contraction mapping* (or a contraction mapping with respect to the metric d) if there exists $0 \leq \lambda < 1$ such that

$$d(F(x), F(y)) \leq \lambda d(x, y) \text{ for all } x, y \in \mathcal{X}.$$

Theorem 2 (**Banach fixed point theorem**). *Let $(\mathcal{X}, d)$ be a complete metric space and $F : \mathcal{X} \to \mathcal{X}$ a d-contraction mapping. Then F has a unique fixed point x^*. Moreover, for all $x_0 \in \mathcal{X}$ the sequence of iterations $x_n := F^n(x_0)$ converges to x^*.*

Proof. Let $0 \leq \lambda < 1$ be the constant involved in the definition of contraction mapping. We begin by showing uniqueness. Let $x, y \in \mathcal{X}$ be such that $F(x) = x$ and $F(y) = y$. Then $d(x, y) = d(F(x), F(y)) \leq \lambda d(x, y)$. Since $d(x, y) \geq 0$ and $\lambda < 1$, it follows that $d(x, y) = 0$, so $x = y$ by identity of indiscernibles.

Now, for any $x_0 \in \mathcal{X}$, we show that the sequence $x_n := F^n(x_0)$ converges to a fixed point of F. First observe that F is continuous, as $d(x, y) < \epsilon/\lambda$ implies $d(F(x), F(y)) < \epsilon$. We also have that $d(x_{n+1}, x_{n+2}) = d(F(x_n), F(x_{n+1})) \leq \lambda d(x_n, x_{n+1})$, and thus (by induction) $d(x_{n+1}, x_n) \leq \lambda^n d(x_1, x_0)$. Therefore, for $n < m$,

$$d(x_m, x_n) \leq \sum_{j=0}^{m-n-1} d(x_{n+j+1}, x_{n+j}) \leq \sum_{j=0}^{m-n-1} \lambda^{j+n} d(x_1, x_0) \leq \frac{\lambda^n - \lambda^m}{1-\lambda} d(x_1, x_0) \leq \lambda^n \frac{d(x_1, x_0)}{1-\lambda},$$

so that (x_n) is a Cauchy sequence. Since $\mathcal{X}$ is complete, it follows that (x_n) converges to some limit x^*. But then x^* is a fixed point of T, since

$$F(x^*) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = x^*,$$

where the second equality is justified by continuity of F. □

For the remainder of this section we briefly discuss the case where L is a (total) bounded operator in a Banach space, and for which there are well-known results.

Proposition 2.1.5 (**The exponential operator**). *Let $\mathcal{X}$ be a Banach space and $L : \mathcal{X} \to \mathcal{X}$ be a bounded operator.*

1. *The series $\sum_{n=0}^{\infty} \frac{1}{n!} L^n$ converges in the operator norm to a bounded operator which we denote by e^L; moreover, we have $\|e^L\| \leq e^{\|L\|}$.*
2. *For $g \in \mathcal{X}$, and $t \in \mathbb{T}$, define $u(t) = e^{tL} g$ so that $u \in C(\mathbb{T}, \mathcal{X})$; then also $u \in C^1(\mathbb{T}, \mathcal{X})$ and*

$$\frac{du}{dt}(t) = Lu(t), \quad \text{for all } t \in \mathbb{T}.$$

Proof. For a proof, see for example [Paz83, Section 1.1]. □

We can use the contraction mapping principle to prove that $\Phi_{L,g}$ has a fixed point. For $T \in \mathbb{T}$, we consider the space $C([0, T], \mathcal{X})$, which is a Banach space with norm

$$\|u\|_{C([0,T],\mathcal{X})} = \sup_{0 \leq t \leq T} \|u(t)\|_{\mathcal{X}}.$$

We can then consider the restriction of $\Phi_{L,g}$ to $C([0, T], \mathcal{X})$, given by

$$\begin{aligned} \Phi_{L,g,T} : \quad C([0,T],\mathcal{X}) &\to C([0,T],\mathcal{X}) \\ u(t) &\mapsto g + \textstyle\int_0^t Lu(s)ds. \end{aligned} \tag{2.4}$$

Even though $\Phi_{L,g,T}$ may not be contracting, we shall see that its nth-iterates $\Phi^n_{L,g,T}$ will be for large enough n. Before we introduce the main result, we give a technical lemma.

Lemma 2.1.6 (**Contraction inequalities in a Banach space**). *Let $\mathcal{X}$ be a Banach space and $L : \mathcal{X} \to \mathcal{X}$ be a bounded operator with norm C. Then for any $g \in \mathcal{X}$, $T \in \mathbb{T}$, $n \in \mathbb{N}$ and $u, v \in C([0,T], \mathcal{X})$, we have*

$$d(\Phi^n_{L,g,T}(u), \Phi^n_{L,g,T}(v)) \leq \frac{(CT)^n}{n!} d(u,v). \tag{2.5}$$

Proof. By induction on n.

Base step with $n = 0$: trivial, the inequality just becomes $d(u,v) \leq d(u,v)$.

Induction step: Assume (2.5) holds for some n and all $T \in \mathbb{T}$. Then

$$\begin{aligned}
d(\Phi^{n+1}_{L,g,T}(u), \Phi^{n+1}_{L,g,T}(v)) &= \|\Phi^{n+1}_{L,g,T}(u) - \Phi^{n+1}_{L,g,T}(v)\|_{C([0,T],\mathcal{X})} && \text{(2.6a)}\\
&= \sup_{0\leq t\leq T} \left\| (\Phi^{n+1}_{L,g,T}(u) - \Phi^{n+1}_{L,g,T}(v)) \right\| && \text{(2.6b)}\\
&= \sup_{0\leq t\leq T} \left\| \int_0^t L(\Phi^n_{L,g,T}(u)) - L(\Phi^n_{L,g,T}(v)) ds \right\| && \text{(2.6c)}\\
&\leq \sup_{0\leq t\leq T} \int_0^t \left\| L(\Phi^n_{L,g,T}(u) - \Phi^n_{L,g,T}(v)) \right\| ds && \text{(2.6d)}\\
&\leq \sup_{0\leq t\leq T} \int_0^t C\|\Phi^n_{L,g,T}(u) - \Phi^n_{L,g,T}(v)\| ds && \text{(2.6e)}\\
&\leq \sup_{0\leq t\leq T} \int_0^t C\frac{(Cs)^n}{n!} d(u,v) ds && \text{(2.6f)}\\
&= \sup_{0\leq t\leq T} \frac{(Ct)^{n+1}}{(n+1)!} d(u,v) = \frac{(CT)^{n+1}}{n!} d(u,v), && \text{(2.6g)}
\end{aligned}$$

where (2.6c) is justified by writing $\Phi^{n+1}_{L,g,T}(u) = \Phi_{L,g,T}(\Phi^n_{L,g,T}(u))$ and applying equation (2.4), (2.6d) by majorizing the integral and by linearity of L, (2.6e) by boundedness of L, (2.6f) by converting to distance and using the induction hypothesis and (2.6g) by simply computing the integral. We thus obtain the desired result. □

Theorem 3 (**Solution to the linear evolution problem with a bounded operator**). *Let $\mathcal{X}$ be a Banach space and $L : \mathcal{X} \to \mathcal{X}$ be a bounded operator. Then for any $g \in \mathcal{X}$ the operator $\Phi_{L,g}$ given by Equation* (2.3) *has a unique fixed point; moreover, this fixed point is given by $u_g(t) = e^{tL}g$.*

Proof. This is an immediate consequence of the results shown in [Paz83]; however, we shall present a different proof that invokes the Banach fixed point theorem (Theorem 2). Let C be the norm of L. For $T \in \mathbb{T}$, consider the space $C([0,T], \mathcal{X})$ as before. From Lemma 2.1.6 we see that, for all n sufficiently large (namely for n such that $(CT)^n < n!$), the nth-iterate $\Phi^n_{L,g,T}$ is a d-contraction mapping. Thus, by the contraction mapping principle (Theorem 2), it follows that $\Phi^n_{L,g,T}$ has a unique fixed point, which we denote by u_n. By the same principle we can assume that $\Phi^{n+1}_{L,g,T}$ and $\Phi^{n(n+1)}_{L,g,T}$ have unique fixed points u_{n+1} and $u_{n(n+1)}$. Now observe that

$$\Phi^{n(n+1)}_{L,g,T}(u_n) = \underbrace{\Phi^n_{L,g,T} \circ \ldots \circ \Phi^n_{L,g,T}}_{n+1 \text{ times}}(u_n) = u_n;$$

by uniqueness of fixed points we conclude that $u_n = u_{n(n+1)}$; similarly we conclude that $u_{n+1} = u_{n(n+1)}$ and thus $u_n = u_{n+1}$. Now

$$u_n = u_{n+1} = \Phi^{n+1}_{L,g,T}(u_{n+1}) = \Phi_{L,g,T}(\Phi^n_{L,g,T}(u_{n+1})) = \Phi_{L,g,T}(\Phi^n_{L,g,T}(u_n)) = \Phi_{L,g,T}(u_n),$$

so u_n is also a fixed point of $\Phi_{L,g,T}$. Since any fixed point of $\Phi_{L,g,T}$ must also be a fixed point of its nth-iterate $\Phi^n_{L,g,T}$ it follows that this is the (finite time) unique fixed point of $\Phi_{L,g,T}$ and so we can denote it by $u_{g,T}$.

Next we study the change in $T \in \mathbb{T}$. Observe that, if $u_{g,T}$ is a fixed point of $\Phi_{L,g,T}$ and $T' \leq T$, then the restriction of $u_{g,T}$ to $[0,T']$ must be a fixed point of $\Phi_{L,g,T'}$ (this is immediate from equation (2.4)). Again by uniqueness, it follows that such restriction must be the fixed point $u_{g,T'}$. Thus we can define the family $\{u_{g,T}\}_{T\in\mathbb{T}}$ of finite time fixed points, where each $u_{g,T}$ is defined in $[0,T]$ and is the restriction of $u_{g,T'}$ for all $T' \geq T$. If we take the limit, we obtain $u_g \in C(\mathbb{T}, \mathcal{X})$ given by $u_g(t) = u_{g,T}(t)$ for any $T \geq t$.

The restriction of the stream u_g to $[0,T]$ is the unique fixed point of $\Phi_{L,g,T}$ for all $T \in \mathbb{T}$, and it is immediate from equation (2.4) that for any function u,

$$\begin{gathered} u \text{ is a fixed point of } \Phi_{L,g} \text{ if and only if} \\ \text{for all } T \in \mathbb{T}, \text{ the restriction } u \restriction_{[0,T]} \text{ is a fixed point of } \Phi_{L,g,T}; \end{gathered}$$

thus we conclude that u_g is the unique fixed point of $\Phi_{L,g}$.

Finally, from Proposition 2.1.5 we see that $t \mapsto e^{tL}g$ defines a solution to the linear evolution problem (2.1) and thus it must be a fixed point of $\Phi_{L,g}$. Therefore it follows that $u_g(t) = e^{tL}g$, which concludes the proof. □

2.2 Fréchet Spaces

For the rest of the chapter we will focus on those linear operators which are not bounded: either they are unbounded, or they operate on a space $\mathcal{X}$ which is not a Banach space. An important family of examples consists of differential operators (such as $L = \partial_x$ with corange $C([0,1])$, seen in Example 2.1.2). In this section we begin by attempting to define a natural space $\mathcal{X}$ for our class of streams $C(\mathbb{T}, \mathcal{X})$, which will lead us into the study of Fréchet spaces.

Some of the properties that we desire of our space $\mathcal{X}$ are that it must be a complete metric space and it must contain only infinitely differentiable functions, since the operators we wish to consider involve taking derivatives. We will see that the space of infinitely differentiable functions will serve as a good candidate for our investigation. This space does not have a norm, however there exist families of *pseudonorms* which induce a complete metric on it. In this way we are led to the concept of Fréchet space, to which we now turn. We refer the reader to [RS80, Chapter V], where a detailed exposition of Fréchet spaces can be found.

Definition 2.2.1 (Pseudonorm). Let $\mathcal{X}$ be a vector space. A *pseudonorm* (sometimes called *seminorm*) is a function $\|\cdot\| : \mathcal{X} \to \mathbb{R}$ which is:

- positive semidefinite, that is, $\|0\| = 0$ and for all $x \in \mathcal{X}\backslash\{0\}$ we have $\|x\| \geq 0$;
- scalable, that is, for all scalars c and $x \in \mathcal{X}$ we have $\|cx\| = |c|\|x\|$;
- subadditive, that is, for all $x, y \in \mathcal{X}$ we have $\|x + y\| \leq \|x\| + \|y\|$.

Recall that a *norm* is a pseudonorm which is also:

- positive definite, that is, $\|0\| = 0$ and for all $x \in \mathcal{X}\backslash\{0\}$ we have $\|x\| > 0$ (this condition replaces positive semidefiniteness).

Example 2.2.2. Consider the space $C(\mathbb{R})$ of all continuous functions of type $\mathbb{R} \to \mathbb{R}$. Then, we can define a family of pseudonorms $\|\cdot\|_n$ (indexed by $n \in \mathbb{N}$) given by

$$\|f\|_n = \sup_{-n \leq x \leq n} |f(x)|. \tag{2.7}$$

Observe that, although no pseudonorm is a norm (that is, for each n there exists $f \in C(\mathbb{R})$ such that $f \neq 0$ but $\|f\|_n = 0$), we have that $\|f\|_n = 0$ for all n implies $f = 0$.

Example 2.2.3. Consider the space $C^\infty([0,1])$ of all functions of type $[0,1] \to \mathbb{R}$ which are infinitely differentiable. For each $k \in \mathbb{N}$ we can define a pseudonorm given by

$$\|f\|_k = \sup_{x \in [0,1]} \left| f^{(k)}(x) \right|. \tag{2.8}$$

As in the previous example, we also have that $f = 0$ iff $\|f\|_n = 0$ for all $n \in \mathbb{N}$.

Example 2.2.4. Consider the space $C^\infty(\mathbb{R})$ of infinitely differentiable real functions. For each $k, n \in \mathbb{N}$ we can define a pseudonorm $\|\cdot\|_{n,k}$ given by

$$\|f\|_{n,k} = \sup_{-n \leq x \leq n} \left| f^{(k)}(x) \right|. \tag{2.9}$$

As in the two previous examples, we have that $f = 0$ iff $\|f\|_{n,k} = 0$ for all $k, n \in \mathbb{N}$.

The previous examples motivate us to establish the following notion.

Definition 2.2.5 (**Point separability**)**.** Let $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ be a family of pseudonorms in a vector space $\mathcal{X}$. We say that the family *separates points* if

$$\|x\|_\alpha = 0 \text{ for all } \alpha \in A \text{ implies } x = 0.$$

In order to be able to define a Fréchet space, we need to recall the notions of topology induced by a family of pseudonorms and completeness with respect to a family of pseudonorms.

Definition 2.2.6 (**Induced topology**)**.** Let $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ be a family of pseudonorms in a vector space $\mathcal{X}$. The *topology induced* by $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ is the topology generated by basic sets of the form $N_{A_0,\epsilon,x}$, for A_0 a finite subset of A, $\epsilon > 0$, $x \in \mathcal{X}$, where

$$N_{A_0,\epsilon,x} = \{y \in \mathcal{X} : \quad \|x - y\|_\alpha < \epsilon, \alpha \in A_0\}.$$

Definition 2.2.7 (**Complete space**)**.** Let $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ be a family of pseudonorms in a vector space $\mathcal{X}$. We say that $\mathcal{X}$ is *complete* with respect to $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ if, for all sequences (x_n) such that (x_n) is Cauchy with respect to each pseudonorm $\|\cdot\|_\alpha$, there exists $x \in \mathcal{X}$ such that (x_n) converges to x with respect to each pseudonorm $\|\cdot\|_\alpha$.

Definition 2.2.8 (**Fréchet space**)**.** A *Fréchet space* is a topological vector space $\mathcal{X}$ with a family of pseudonorms $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ with the following properties:

- A is countable;
- $\{\|\cdot\|_\alpha\}_{\alpha \in A}$ separates points;

- the topology on $\mathcal{X}$ is induced by $\{\|\cdot\|_\alpha\}_{\alpha\in A}$;
- $\mathcal{X}$ is complete with respect to $\{\|\cdot\|_\alpha\}_{\alpha\in A}$.

Example 2.2.9. Examples 2.2.2, 2.2.3 and 2.2.4 are Fréchet spaces with their correspondent families of pseudonorms and induced topologies.

Example 2.2.10 (**Schwarz space**)**.** Consider the *Schwarz space*, denoted by $\mathcal{S}(\mathbb{R})$, of rapidly decreasing functions, that is, infinitely differentiable real functions f for which

$$\|f\|_{a,b} = \sup_{x\in\mathbb{R}} |x^a f^{(b)}(x)| < \infty, \tag{2.10}$$

for all $a, b \in \mathbb{N}$. Then $(\|\cdot\|_{a,b})$ is a countable family of pseudonorms whose induced topology in $\mathcal{S}(\mathbb{R})$ turns it into a Fréchet space.

Example 2.2.11. The space $C(\mathbb{T}, C^\infty(\mathbb{R}))$ of continuous streams over $C^\infty(\mathbb{R})$ is also a Fréchet space, with the countable family of pseudonorms given by

$$\|u\|_{T,n,k} = \sup_{0\le t\le T} \sup_{-n\le x\le n} \left|\frac{\partial^k u}{\partial x^k}(t,x)\right|, \tag{2.11}$$

for $T, n, k \in \mathbb{N}$.

We can see that the family of pseudonorms in Example 2.2.11 is closely related to the family in Example 2.2.4. In fact, this illustrates a useful property of Fréchet spaces; in general, the space of continuous streams over a Fréchet space is itself a Fréchet space. In other words, Fréchet spaces work well with the operation of taking continuous streams.

Proposition 2.2.12 (**New Fréchet spaces from old**)**.** *If $\mathcal{X}$ is a Fréchet space with a countable family of pseudonorms $\{\|\cdot\|_\alpha\}_{\alpha\in A}$, then so is $C(\mathbb{T}, \mathcal{X})$ with the countable family of pseudonorms $\{\|\cdot\|_{T,\alpha}\}_{T\in\mathbb{N},\alpha\in A}$, where*

$$\|u\|_{T,\alpha} = \sup_{0\le t\le T} \|u(t)\|_\alpha.$$

The next step is to use the family of pseudonorms to devise a metric on the Fréchet space. First observe that, if the Fréchet space is described by a finite family of pseudonorms $\{\|\cdot\|_1, \ldots, \|\cdot\|_n\}$, then we can easily devise a metric by adding all pseudonorms:

$$d(x,y) = \sum_{i=1}^{n} \|x-y\|_i.$$

However, when we have a countably infinite family of pseudonorms, we cannot, in general, use a summation over all pseudonorms, since we need a convergent series. To overcome this obstacle we use two techniques: enforce a bound on each term on the summation and introduce a summable family of weights.

Proposition 2.2.13 (**Metric from family of pseudonorms**)**.** *Let $\mathcal{X}$ be a Fréchet space and $\{\|\cdot\|_\alpha\}_{\alpha\in A}$ a corresponding family of pseudonorms. Let $\gamma : \mathbb{R}^{\ge 0} \to [0,1]$ be a continuous function which is also positive definite, increasing and subadditive, that is,*

- *$\gamma(0) = 0$ and for all $t \in \mathbb{R}^+$ we have $0 < \gamma(t) \le 1$;*
- *for all $t_1, t_2 \in \mathbb{R}_0^+$ such that $t_1 \le t_2$ we have $\gamma(t_1) \le \gamma(t_2)$;*

- *for all* $t_1, t_2 \in \mathbb{R}_0^+$ *we have* $\gamma(t_1 + t_2) \leq \gamma(t_1) + \gamma(t_2)$.

Let $\{w_\alpha\}_{\alpha \in A}$ *be a summable family of positive weights, that is,* $\sum_{\alpha \in A} w_\alpha < \infty$. *Then we can define a metric on* $\mathcal{X}$ *by*

$$d(x, y) = \sum_{\alpha \in A} w_\alpha \gamma(\|x - y\|_\alpha). \tag{2.12}$$

Moreover, this metric induces the same topology over $\mathcal{X}$ *and* $\mathcal{X}$ *is complete under it.*

Proof. We begin by showing that d defines a metric:

Finiteness: for any $x, y \in \mathcal{X}$ we have

$$d(x, y) = \sum_{\alpha \in A} w_\alpha \gamma(\|x - y\|_\alpha) \leq \sum_{\alpha \in A} w_\alpha < \infty,$$

where we have used boundedness of γ and summability of the family of weights.

Positive definiteness: for any $x, y \in \mathcal{X}$, if $x = y$ then $x - y = 0$ and so

$$d(x, y) = \sum_{\alpha \in A} w_\alpha \gamma(\|0\|_\alpha) = \sum_{\alpha \in A} 0 = 0,$$

where the second equality is justified by positive semidefiniteness of the pseudonorms and positive definiteness of γ. Moreover, if $x \neq y$ then there is $\alpha \in A$ such that $\|x - y\|_\alpha > 0$ (by positive semidefiniteness of the pseudonorms and point separability of the family of pseudonorms) and thus

$$d(x, y) = \sum_{\alpha \in A} w_\alpha \gamma(\|x - y\|_\alpha) \geq w_\alpha \gamma(\|x - y\|_\alpha) > 0,$$

where the last step is justified by positivity of the weights and positive definiteness of γ.

Symmetry: for any $x, y \in \mathcal{X}$ we have

$$d(x, y) = \sum_{\alpha \in A} w_\alpha \gamma(\|x - y\|_\alpha) = \sum_{\alpha \in A} w_\alpha \gamma(\|y - x\|_\alpha) = d(y, x),$$

where the second step is justified by symmetry of the pseudonorms.

Triangle inequality: for any $x, y, z \in \mathcal{X}$ we have

$$\begin{aligned} d(x, z) &= \sum_{\alpha \in A} w_\alpha \gamma(\|x - z\|_\alpha) && \text{(2.13a)} \\ &\leq \sum_{\alpha \in A} w_\alpha \gamma(\|x - y\|_\alpha + \|y - z\|_\alpha) && \text{(2.13b)} \\ &\leq \sum_{\alpha \in A} w_\alpha (\gamma(\|x - y\|_\alpha) + \gamma(\|y - z\|_\alpha)) && \text{(2.13c)} \\ &= d(x, y) + d(y, z), && \text{(2.13d)} \end{aligned}$$

where (2.13b) is justified by subadditivity of the pseudonorms and (2.13c) is justified by subadditivity of γ.

The next step is to show equivalence between the topology generated by the pseudonorms (denoted by $\mathcal{T}_\rho$) and the topology induced by the metric (denoted by $\mathcal{T}_d$).

$\mathcal{T}_\rho \subseteq \mathcal{T}_d$: Let A_0 be a finite subset of A, $\epsilon > 0$ and $x \in \mathcal{X}$. Let $N = N_{A_0, \epsilon, x}$ be a basic open set

in $\mathcal{T}_\rho$ and $y \in N$; in this way, $\|x-y\|_\alpha < \epsilon$ for $\alpha \in A_0$. Let $\delta = \max_{\alpha \in A_0} \|x-y\|_\alpha$, so that $\delta < \epsilon$, and let $w = \min_{\alpha \in A_0} w_\alpha$. Now take $B = B_{w\gamma(\epsilon-\delta)}(y)$, that is, the ball centered at y with radius $w\gamma(\epsilon-\delta)$. Clearly $y \in B$. To show that $B \subseteq N$, take $z \in B$, so that $d(y,z) < w\gamma(\epsilon-\delta)$. In particular, for all $\alpha \in A_0$, we have $w_\alpha \gamma(\|y-z\|_\alpha) \leq d(y,z) < w\gamma(\epsilon-\delta)$, so that $\gamma(\|y-z\|_\alpha) < \dfrac{w}{w_\alpha}\gamma(\epsilon-\delta) \leq \gamma(\epsilon-\delta)$ and thus $\|y-z\|_\alpha < \epsilon - \delta$ by monotonicity of γ. Finally, we conclude that, for all $\alpha \in A_0$, we have by subadditivity of the pseudonorms that $\|x-z\|_\alpha \leq \|x-y\|_\alpha + \|y-z\|_\alpha < \delta + \epsilon - \delta = \epsilon$, which implies that $z \in N$.

$\mathcal{T}_d \subseteq \mathcal{T}_\rho$: Let $B = B_\epsilon(x)$ be a basic open set in $\mathcal{T}_d$ and $y \in B$, so that $d(x,y) < \epsilon$. Let $\delta = d(x,y)$, let A_0 be a sufficiently large, yet finite, subset of A such that $\sum_{\alpha \in A \setminus A_0} w_\alpha < \dfrac{\epsilon-\delta}{2}$ and let $w = \sum_{\alpha \in A_0} w_\alpha$. Let also $\beta \in \mathbb{R}^+$ be such that $\gamma(\beta) \leq \dfrac{\epsilon-\delta}{2w}$, which is possible by continuity of γ. Now take $N = N_{A_0,\beta,y}$. Clearly $y \in N$. To show that $N \subseteq B$, take $z \in N$, so that $\|y-z\|_\alpha < \beta$ for all $\alpha \in A_0$ and thus $\gamma(\|y-z\|_\alpha) < \dfrac{\epsilon-\delta}{2w}$ by monotonicity of γ. Therefore

$$d(y,z) = \sum_{\alpha \in A} w_\alpha \gamma(\|y-z\|_\alpha) = \sum_{\alpha \in A_0} w_\alpha \gamma(\|y-z\|_\alpha) + \sum_{\alpha \in A \setminus A_0} w_\alpha \gamma(\|y-z\|_\alpha) \tag{2.14a}$$

$$< \sum_{\alpha \in A_0} \frac{w_\alpha}{w} \frac{\epsilon-\delta}{2} + \sum_{\alpha \in A \setminus A_0} w_\alpha \tag{2.14b}$$

$$\leq \frac{\epsilon-\delta}{2} + \frac{\epsilon-\delta}{2} = \epsilon - \delta, \tag{2.14c}$$

where (2.14b) is justified by boundedness of γ and the above paragraph, and (2.14c) is justified by the above paragraph. Finally, we conclude by triangle inequality that $d(x,z) \leq d(x,y) + d(y,z) < \delta + \epsilon - \delta = \epsilon$, which implies that $z \in B$.

The only step that remains is showing that $\mathcal{X}$ is complete under the metric d. Let (x_n) be a Cauchy sequence with respect to the metric d. That (x_n) is also a Cauchy sequence with respect to each pseudonorm $\|\cdot\|_\alpha$ follows from the inequalities $0 \leq w_\alpha \gamma(\|x-y\|_\alpha) \leq d(x,y)$ and monotonicity and positive definiteness of γ. Since $\mathcal{X}$ is complete with respect to $\{\|\cdot\|_\alpha\}_{\alpha \in A}$, it follows that there exists $x \in \mathcal{X}$ such that (x_n) converges to x in each pseudonorm $\|\cdot\|_\alpha$. Then (x_n) converges to x in the topology induced by the family of pseudonorms, which means it converges with respect to the metric d since it induces the same topology. □

Common choices for the function γ are $\gamma_1(t) = \frac{t}{1+t}$ and $\gamma_2(t) = \min(t,1)$.

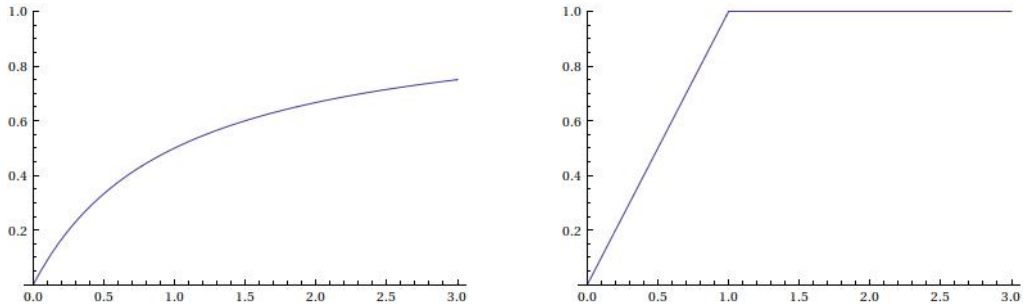


Figure 2.3: Plot of $\gamma_1(t) = \frac{t}{1+t}$ (left) and $\gamma_2(t) = \min(t, 1)$ (right).

Example 2.2.14. Recall the space of infinitely differentiable functions $f : [0, 1] \to \mathbb{R}$ mentioned in Example 2.2.3. This space is complete under the metric

$$d(f, g) = \sum_{k=0}^{\infty} 2^{-k} \frac{\|f - g\|_k}{1 + \|f - g\|_k}. \tag{2.15}$$

In later chapters we shall use a specific choice of metric induced by the pseudonorms in which $A = \mathbb{N}^+$, $w_n = 2^{-n}$ and $\gamma(t) = \min(t, 1)$. It will be useful to have a result relating bounds on the pseudonorms with bounds on the metric, as we now state and prove.

Proposition 2.2.15 (**Bounds on the pseudonorms vs bounds on the metric**). *Let $\mathcal{X}$ be a Fréchet space with pseudonorms $\|\cdot\|_n$, $n \in \mathbb{N}^+$. Let d be the metric on $\mathcal{X}$ given by*

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min(\|x - y\|_n, 1). \tag{2.16}$$

1. *Let $0 < \epsilon < 1$ and $M \in \mathbb{N}$. Then, for any $\delta \leq \epsilon 2^{-M}$ and $x, y \in \mathcal{X}$, one has*

   $$\text{if } d(x, y) < \delta, \text{ then } \|x - y\|_n < \epsilon \text{ for } n = 1, \ldots, M.$$

2. *Let $0 < \epsilon < 1$. Then for any $\delta \leq \epsilon/2$ and $M \in \mathbb{N}$ such that $2^{-M} \leq \epsilon/2$ and $x, y \in \mathcal{X}$, one has*

   $$\text{if } \|x - y\|_n < \delta \text{ for } n = 1, \ldots, M, \text{ then } d(x, y) < \epsilon.$$

Proof. To prove the first claim, take ϵ, M, δ as in the assumptions and $x, y \in \mathcal{X}$ with $d(x, y) < \delta$. Since $d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min(\|x - y\|_n, 1)$ it follows that, for all $n \in \mathbb{N}$, we have that

$$2^{-n} \min(\|x - y\|_n, 1) < \delta, \text{ that is, } \min(\|x - y\|_n, 1) < \delta 2^n.$$

By using the bound on δ and considering only the values of n between 1 and M, we see that $\min(\|x - y\|_n, 1) < \delta 2^n < \epsilon 2^{-M} 2^n \leq \epsilon$. Since $\epsilon < 1$ this further implies that $\|x - y\|_n < \epsilon$ for $n = 1, \ldots, M$, as we wanted to prove.

To prove the second claim, take ϵ, δ, M as in the assumptions and $x, y \in \mathcal{X}$ with $\|x - y\|_n < \delta$ for $n = 1, \ldots, M$. By splitting the sum in (2.16) we get

$$d(x, y) = \sum_{n=1}^{\infty} 2^{-n} \min(\|x - y\|_n, 1)$$

$$
\begin{aligned}
&= \sum_{n=1}^{M} 2^{-n} \min(\|x-y\|_n, 1) + \sum_{n=M+1}^{\infty} 2^{-n} \min(\|x-y\|_n, 1) \\
&\leq \sum_{n=1}^{M} 2^{-n} \|x-y\|_n + \sum_{n=M+1}^{\infty} 2^{-n} \\
&< \sum_{n=1}^{M} 2^{-n} \delta + 2^{-M} \\
&< \delta + 2^{-M} < \epsilon/2 + \epsilon/2 = \epsilon,
\end{aligned}
$$

as we wanted to prove. □

2.3 Iterating scheme

For the rest of this chapter we shall assume that $\mathcal{X}$ (and thus, also $C(\mathbb{T}, \mathcal{X})$) always denotes a Fréchet space. We recall that, given a linear operator L and an initial condition g defining (2.1), we can consider the operator $\Phi_{L,g} : C(\mathbb{T}, \mathcal{X}) \to C(\mathbb{T}, \mathcal{X})$ given by

$$\Phi_{L,g}(u)(t) = g + \int_0^t Lu(s)ds. \tag{2.17}$$

We know that solutions to (2.1) correspond to fixed points of $\Phi_{L,g}$. One important advantage of choosing a linear operator L is made obvious in the following result.

Lemma 2.3.1. *Consider the problem* (2.1)*, where g is in the intersection $\bigcap_{n\in\mathbb{N}} D(L^n)$. Then the sequence u_k obtained by iterating $\Phi_{L,g}$ (defined in* (2.17)*) with $u_0(t) = g$ is given by*

$$u_k(t) = g + L(g)t + \ldots + L^k(g)\frac{t^k}{k!} \tag{2.18}$$

Proof. By induction.

Base step: When $k = 0$, we have

$$u_0(t) = g = \sum_{i=0}^{0} L^i(g)\frac{t^i}{i!}.$$

Induction step: Assume (2.18) holds for k. Then

$$
\begin{aligned}
u_{k+1}(t) &= \Phi_{L,g}(u_k)(t) && \text{(2.19a)} \\
&= g + \int_{t_0}^{t} L(u_k)(s)ds && \text{(2.19b)} \\
&= g + \int_{t_0}^{t} L\left(\sum_{i=0}^{k} L^i(g)\frac{s^i}{i!}\right) ds && \text{(2.19c)} \\
&= g + \int_{t_0}^{t} \sum_{i=0}^{k} L(L^i(g))\frac{s^i}{i!} ds && \text{(2.19d)}
\end{aligned}
$$

$$= g + \sum_{i=0}^{k} \int_{t_0}^{t} L^{i+1}(g) \frac{s^i}{i!} ds \tag{2.19e}$$

$$= g + \sum_{i=0}^{k} L^{i+1}(g) \frac{t^{i+1}}{(i+1)!} = \sum_{i=0}^{k+1} L^i(g) \frac{t^i}{i!}, \tag{2.19f}$$

where (2.19c) is justified by induction hypothesis and (2.19d) by linearity of L. We obtain the desired result. □

We want to find suitable conditions under which the sequence u_n defined iteratively by $u_{n+1} = \Phi_{L,g}(u_n)$ (with some initial input u_0) is convergent. The most simple case is that of nilpotent streams.

Definition 2.3.2 (**Nilpotency**)**.** We say that a stream $g \in \mathcal{X}$ is L-nilpotent if there is some $k \in \mathbb{N}$ such that $L^k g = 0$.

Theorem 4 (**Existence of fixed points for nilpotent initial condition**)**.** *Suppose L is a linear differential operator. Consider the problem* (2.1)*, where g is infinitely differentiable. Suppose also that g is nilpotent, and in particular let $k \in \mathbb{N}$ be such that $L^{k+1} g = 0$. Then* (2.1) *has (at least) one solution given by*

$$u(t) = u_k(t) = \sum_{i=0}^{k} L^i(g) \frac{t^i}{i!}. \tag{2.20}$$

Proof. Observe that

$$\Phi_{L,g}(u_k)(t) = u_{k+1}(t) \overset{1}{=} \sum_{i=0}^{k+1} L^i(g) \frac{t^i}{i!} \overset{2}{=} \sum_{i=0}^{k} L^i(g) \frac{t^i}{i!} \overset{3}{=} u_k(t),$$

where (1) and (3) are justified by Lemma 2.3.1 and (2) by nilpotency of g. Hence u_k is a fixed point of $\Phi_{L,g}$, and thus it is a solution to (2.1). □

Example 2.3.3. Consider $\mathcal{X} = C^\infty(\mathbb{R})$, $L = \partial_x$ and $g(x) = x^2 + 2x$.

Then

$$L(g)(x) = 2x + 2; \qquad L^2(g)(x) = 2; \qquad L^3(g) = 0.$$

We conclude, using Theorem 4, that (2.1) has a solution given by $u = x^2 + 2x + (2x+2)t + t^2 = (x+t)^2 + 2(x+t)$. This solution could also be achieved by iteration on $\Phi_{L,g}$. □

Example 2.3.4. Consider $\mathcal{X} = C^\infty(\mathbb{R}^2)$, $L = \Delta = \partial_x^2 + \partial_y^2$ (the Laplacian operator) and $g(x,y) = x^4 + xy^2 + y^3$.

Then

$$L(g)(x,y) = 12x^3 + 2x + 6y; \qquad L^2(g)(x,y) = 72x; \qquad L^3(g) = 0.$$

We conclude, using Theorem 4, that (2.1) has a solution given by $u = x^4 + xy^2 + y^3 + (12x^3 + 2x + 6y)t + 36xt^2$. This solution could also be achieved by iteration on $\Phi_{L,g}$. □

Definition 2.3.5 (**Absolute convergence**). Let (a_k) be a sequence in $\mathcal{X}$. We say that the series $\sum_{k_0}^{\infty} a_k$ is absolutely convergent if for all $\alpha \in A$, $\sum_{k_0}^{\infty} \|a_k\|_\alpha$ is a convergent series.

Proposition 2.3.6. *Any absolutely convergent series $\sum_{k_0}^{\infty} a_k$ is convergent.*

Theorem 5 (**Existence of fixed points for absolutely convergent power series**). *Let $\mathcal{X}$ be a Fréchet space, and suppose $L : \mathcal{X} \to \mathcal{X}$ is total and continuous. Consider the problem (2.1), where $g \in \mathcal{X}$. Suppose also that there exists some $T \in \mathbb{T}$ such that the power series $\sum_k a_k t^k$, with $a_k = \frac{L^k(g)}{k!}$ and $t \in [0,T]$, is absolutely convergent. Then the sequence (u_n) in $C([0,T],\mathcal{X})$ obtained by iterating $\Phi_{L,g,T}$, given by (2.4), with $u_0(t) = g$, converges in $C([0,T],\mathcal{X})$ to a finite time solution of (2.1).*

Proof. This result probably exists already in some standard textbook; however, we were not able to find a suitable reference, so we just prove it here. We know from Lemma 2.3.1 that $u_n = \sum_{k=0}^{n} a_k t^k$. Since the power series is absolutely convergent, we know that u_n converges to $u := \sum_k a_k t^k$. We shall show that $\frac{du}{dt}$ and Lu are well-defined, coincide, and are given by

$$\frac{du}{dt} = Lu = \sum_{k=0}^{\infty} (k+1) a_{k+1} t^k.$$

The series converges: we prove that $\sum (k+1) a_{k+1} t^k$ is absolutely convergent, by comparing

$$\sum (k+1) \|a_{k+1}\|_\alpha t^k \quad \text{vs} \quad \sum \|a_k\|_\alpha t^k.$$

It is clear that $\limsup \sqrt[k]{(k+1)\|a_{k+1}\|_\alpha} = \limsup \sqrt[k]{\|a_k\|_\alpha}$, so that both power series must have the same radius of convergence. Thus, $\sum (k+1) a_{k+1} t^k$ is absolutely convergent for any $t \in [0,T]$.

$L(u)$ **is well-defined**: simply observe that

$$\begin{aligned} L(u) = L\left(\sum_{k=0}^{\infty} a_k t^k\right) &= L\left(\lim_{K\to\infty} \sum_{k=0}^{K} a_k t^k\right) = \lim_{K\to\infty} L\left(\sum_{k=0}^{K} a_k t^k\right) \\ &= \lim_{K\to\infty} \sum_{k=0}^{K} L(a_k) t^k = \sum_{k=0}^{\infty} (k+1) a_{k+1} t^k, \end{aligned}$$

where we have used continuity of L, linearity of L and definition of a_k. Since the power series is convergent, we conclude that $L(u)$ is well-defined and coincides with the power series.

$\frac{du}{dt}$ **is well-defined**: let $t \in [0,T]$ and consider $s \in [0,T]$ converging to t, $s \to t$. We now look at the expression

$$\frac{u(s) - u(t)}{s-t} = \frac{\sum a_k s^k - \sum a_k t^k}{s-t} = \sum_{k=0}^{\infty} a_k \frac{s^k - t^k}{s-t} = \sum_{k=1}^{\infty} a_k (s^{k-1} + s^{k-2} t + \ldots + t^{k-1}). \tag{2.21}$$

The summation term can be bounded, for any $\alpha \in A$, by

$$\|a_k(s^{k-1} + s^{k-2}t + \ldots + t^{k-1})\|_\alpha \leq \|a_k\|_\alpha(T^{k-1} + T^{k-1} + \ldots + T^{k-1}) = k\|a_k\|_\alpha T^{k-1}.$$

Since the power series $\sum k\|a_k\|_\alpha t^{k-1}$ is convergent for $t \in [0, T]$, we conclude that the last series in (2.21) is absolutely convergent. Since the above bound does not depend on t or s, the series is also uniformly convergent. Thus we may compute

$$\begin{aligned}
\frac{du}{dt}(t) &= \lim_{s\to t} \frac{u(s) - u(t)}{s - t} && \text{(2.22a)}\\
&= \lim_{s\to t} \sum_{k=1}^{\infty} a_k(s^{k-1} + s^{k-2}t + \ldots + t^{k-1}) && \text{(2.22b)}\\
&= \sum_{k=1}^{\infty} \lim_{s\to t} a_k(s^{k-1} + s^{k-2}t + \ldots + t^{k-1}) && \text{(2.22c)}\\
&= \sum_{k=1}^{\infty} a_k k t^{k-1} = \sum_{k=0}^{\infty} (k+1)a_{k+1}t^k, && \text{(2.22d)}
\end{aligned}$$

where (2.22b) is justified by equation (2.21) and (2.22c) by uniform convergence of the series. We conclude that $\frac{du}{dt}$ and Lu are both equal to the desired series.

Finally, since $u(0) = a_0 = g$, the initial condition is also satisfied, and thus u is a finite time solution of (2.1). □

Remark 2.3.7. We introduce here a small discussion on how Theorem 5 generalises the result on Theorem 3 for Banach spaces. Any Banach space can be seen as a Fréchet space and, in Banach spaces, boundedness is equivalent to continuity (this is a standard result seen, for example, in [Rud91]). Moreover any bounded operator can be extended to a total, bounded operator by the Hahn-Banach Theorem (Theorem 1), and therefore the operator L considered in the statement of Theorem 3 can also be used in Theorem 5.

At a first glance, it seems then that the condition that L be continuous is too strong for the type of operators that we are considering. In particular, we are interested in differential operators such as $L = \partial_x$ or $L = \partial_{xx}$, and the discussion at the start of this chapter (and Example 2.1.2) may lead us to argue that these operators are discontinuous. However, since we are working in a different space $\mathcal{X}$ (which is a Fréchet space), it turns out that differential operators become continuous! We illustrate this with the following example.

Example 2.3.8. Consider $\mathcal{X} = C^\infty(\mathbb{R})$ with pseudonorms defined in Example 2.2.4 and $L = \partial_x$. We show that $L : \mathcal{X} \to \mathcal{X}$ is a continuous operator. Let (a_m) be a sequence in $\mathcal{X}$ converging to 0. Then, for any $n, k \in \mathbb{N}$, we have that

$$\sup_{-n\leq x\leq n} |a_m^{(k)}(x)| \underset{m\to\infty}{\longrightarrow} 0.$$

In particular, for any compact interval $[-n, n]$, not only does a_m (as a function) converge uniformly to zero, but so do all of its derivatives (as functions of type $\mathbb{R} \to \mathbb{R}$). Therefore, we conclude that $\partial_x a_m$ (and also any of its derivatives) must converge uniformly to zero in compact intervals, which proves that $\partial_x a_m \to 0$ in the Fréchet space $\mathcal{X}$. We thus conclude that $L = \partial_x$ is continuous.

Example 2.3.9. Consider $\mathcal{X} = C^\infty(\mathbb{R}^2)$, $L = \Delta = \partial_x^2 + \partial_y^2$ and $g(x, y) = e^x y + e^{2y}$. Observe that L is a continuous operator on $\mathcal{X}$. We have that

$$L(g)(x,y) = e^x y + 4e^{2y}; \qquad L^2(g)(x,y) = e^x y + 16e^{2y};$$

and, in general,

$$L^k(g)(x,y) = e^x y + 4^k e^{2y}.$$

For all x and y, we have convergence of the power series on $t \in \mathbb{T}$,

$$\sum_{k=0}^{\infty} \frac{e^x y + 4^k e^{2y}}{k!} t^k = e^x y \sum_{k=0}^{\infty} \frac{t^k}{k!} + e^{2y} \sum_{k=0}^{\infty} \frac{4^k t^k}{k!} = y e^x e^t + e^{2y} e^{4t}.$$

Hence, using Theorem 5, we conclude that u_k converges to a solution of (2.1), given by $u(t,x,y) = ye^{x+t} + e^{2y+4t}$.

2.4 Convergence theorems for the transport equation

Theorem 5 of the previous section gives us a criterion of classes of operators and initial conditions for which existence of fixed points can be ensured. However, determining directly whether the power series described in its statement is convergent can be a challenging task. In this section we shall focus on the one-dimensional transport equation; we aim to identify certain classes of operators and initial conditions for which convergence occurs. The properties of interest in the following theorems will rely upon analyticity of the initial condition g or the initial input u_0. This will suggest a parallelism with the Cauchy-Kowalevski theorems, which are local existence and uniqueness results for analytic PDEs (a relatively modern reference is [CH53]).

Theorem 6 (**Cauchy-Kowalevski Theorem, [Cau42, Kow75]**). *Consider the time evolution problem*

$$\begin{cases} \partial_t^k u = F(x,t,\partial_t^j \partial_x^\alpha u), & t \in \mathbb{T}, x \in \mathbb{R}, j < k, \alpha + j \leq k; \\ \partial_t^j u(0,x) = g_j(x), & 0 \leq j < k. \end{cases} \tag{2.23}$$

Suppose that F and g_j are real analytic near the origin. Then (2.23) *has a unique real analytic solution near the origin.*

For the next sections, we will be working with linear evolution problems that satisfy the conditions on the Cauchy-Kowalevski Theorem. Since (local) existence and uniqueness of solutions are already given by that theorem, we shall study convergence of the iterating scheme presented in Section 2.3 to the known solution. In particular, we set $\mathcal{X} = C^\infty(\mathbb{R})$, so that $C(\mathbb{T},\mathcal{X}) = C(\mathbb{T}, C^\infty(\mathbb{R}))$, and $L = \alpha\partial_x$ for some $\alpha \in \mathbb{R}$; let us then drop the subscript L and write our iterating operator as

$$\Phi_g(u)(t,x) = g(x) + \alpha \int_0^t \partial_x u(s,x) ds. \tag{2.24}$$

Take any (arbitrary but fixed) $X \in \mathbb{R}^+$, $T \in \mathbb{T}$. Then, for any $k \in \mathbb{N}$, we have a pseudonorm

$$\|u\|_{T,X,k} = \sup_{\substack{0 \leq t \leq T \\ |x| \leq X}} \left| \frac{\partial^k u}{\partial x^k}(t,x) \right|. \tag{2.25}$$

Observe that we are taking suprema on compact rectangles of the form $[0,T]\times[-X,X]$ (see Figure 2.4). The reason for taking suprema on compact rectangles will be made clear shortly in Theorem 7 (local convergence theorem). We also observe that, for each compact rectangle $\mathbb{X} = [0,T]\times[-X,X]$, we can define the space of *compact continuous streams* $C([0,T], C^\infty(-X,X))$. Clearly, any function

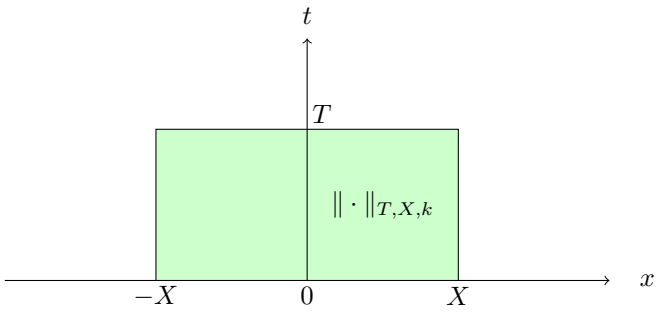


Figure 2.4: Compact rectangles.

in $C(\mathbb{T}, C^\infty(\mathbb{R}))$ can be mapped to a function in $C([0,T], C^\infty(-X,X))$ via the restriction $u \mapsto u \restriction_{\mathbb{X}}$. Moreover, $C([0,T], C^\infty(-X,X))$ can be seen to be a Fréchet space with the family of pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Note that X and T are *fixed* and the indexing is on $k \in \mathbb{N}$.

Finally, observe that the operator $\Phi_g : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ has a restriction $\Phi_g \restriction_{\mathbb{X}}$ to the space $C([0,T], C^\infty(-X,X))$.

Our next step is to prove contraction inequalities, which play an important role in fixed point techniques.

Lemma 2.4.1 (**Contraction inequalities**). *Consider the Fréchet space $C(\mathbb{T}, C^\infty(\mathbb{R}))$ with pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Let $g \in C^\infty(\mathbb{R})$ and $\Phi_g : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by (2.24). Then, for any $u, v \in C(\mathbb{T}, C^\infty(\mathbb{R}))$, any pseudonorm $\|\cdot\|_{T,X,k}$ and any $m \in \mathbb{N}$, we have the following bound:*

$$\|\Phi_g^m(u) - \Phi_g^m(v)\|_{T,X,k} \leq \frac{(|\alpha|T)^m}{m!}\|u - v\|_{T,X,k+m}. \tag{2.26}$$

Proof. By induction on m.

Base step with $m = 0$: trivial.

Induction step: Assume (2.26) holds for any choice of pseudonorm $\|\cdot\|_{T,X,k}$ and some m. Then

$$\begin{aligned}
\|\Phi_g^{m+1}(u) - \Phi_g^{m+1}(v)\|_{T,X,k} &= \sup_{0\leq t\leq T}\sup_{|x|\leq X}\left|\frac{\partial^k}{\partial x^k}(\Phi_g^{m+1}(u) - \Phi_g^{m+1}(v))(t,x)\right| && \text{(2.27a)}\\
&= \sup_{0\leq t\leq T}\sup_{|x|\leq X}\left|\frac{\partial^k}{\partial x^k}\left(\alpha\int_0^t \partial_x(\Phi_g^m(u) - \Phi_g^m(v))(s,x)ds\right)\right| && \text{(2.27b)}\\
&= \sup_{0\leq t\leq T}\sup_{|x|\leq X}\left|\alpha\int_0^t \frac{\partial^{k+1}}{\partial x^{k+1}}(\Phi_g^m(u) - \Phi_g^m(v))(s,x)ds\right| && \text{(2.27c)}\\
&\leq \sup_{0\leq t\leq T}\sup_{|x|\leq X}\left|\alpha\int_0^t \|\Phi_g^m(u) - \Phi_g^m(v)\|_{s,X,k+1}ds\right| && \text{(2.27d)}\\
&\leq \sup_{0\leq t\leq T}\sup_{|x|\leq X}\left|\alpha\int_0^t \frac{(|\alpha|s)^m}{m!}\|u - v\|_{s,X,k+m+1}ds\right| && \text{(2.27e)}\\
&\leq \sup_{0\leq t\leq T}\left|\alpha\int_0^t \frac{(|\alpha|s)^m}{m!}\|u - v\|_{T,X,k+m+1}ds\right| && \text{(2.27f)}
\end{aligned}$$

$$= \sup_{0\leq t\leq T}\left|\alpha\frac{|\alpha|^m t^{m+1}}{(m+1)!}\right| \|u-v\|_{T,X,k+m+1} \tag{2.27g}$$

$$= \frac{(|\alpha|T)^{m+1}}{(m+1)!}\|u-v\|_{T,X,k+m+1}, \tag{2.27h}$$

$$\tag{2.27i}$$

where (2.27b) is justified by writing $\Phi_g^{m+1}(u) = \Phi_g(\Phi_g^m(u))$ and applying equation (2.24), (2.27c) by the Leibniz rule, (2.27d) by majorizing the integral and equation (2.25), (2.27e) by induction hypothesis and (2.27f) by noticing the independence on x and majorizing the pseudonorm. We thus obtain the desired result. □

Let us see how we can use these bounds in a proof.

Theorem 7 (**Local Fréchet space convergence theorem**). *Let $C(\mathbb{T}, C^\infty(\mathbb{R}))$ be the Fréchet space with pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Take an initial input $u_0 \in C(\mathbb{T}, C^\infty(\mathbb{R}))$ and initial condition $g \in C^\infty(\mathbb{R})$. Assume also that $u_0 = 0$ and g is analytic at 0 with some radius of convergence[1] R. Let $\Phi_g : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by (2.24). Then, for any $T, X \in \mathbb{R}^+$ such that $|\alpha|T + X < R$, the sequence (u_m) given by $u_m = \Phi_g^m(u_0)$ converges in the rectangle $\mathbb{X}' = [0,T]\times[-X,X]$ to a fixed point of $\Phi_g \restriction_{\mathbb{X}}$.*

Proof. To facilitate the exposition we introduce the pseudonorms on g given by

$$\|g\|_{X,k} = \sup_{|x|\leq X}\left|\frac{\partial^k g}{\partial x^k}(x)\right|, \quad \text{for} \quad X\in\mathbb{R}^+,\ k\in\mathbb{N}. \tag{2.28}$$

Since g is analytic at 0 with radius of convergence R, there is a sequence of real coefficients (a_j) such that, for all $x \in (-R, R)$,

$$g(x) = \sum_{j=0}^{\infty} a_j x^j.$$

It also follows that $\limsup_{n\to\infty} \sqrt[n]{|a_n|} \leq \frac{1}{R}$ (see Footnote 1). Moreover, we have the following bound, for any $X < R$:

$$\|g\|_{X,k} = \left|\sup_{|x|<X}\sum_{j=0}^{\infty}\frac{(j+k)!}{j!}a_{j+k}x^j\right| \leq \sum_{j=0}^{\infty}\frac{(j+k)!}{j!}|a_{j+k}|X^j. \tag{2.29}$$

Let $T, X \in \mathbb{R}^+$ such that $|\alpha|T + X < R$. We show that (u_m) is a Cauchy sequence with respect to the pseudonorm $\|\cdot\|_{T,X,k}$. First observe that

$$\sum_{m=0}^{\infty}\|u_{m+1}-u_m\|_{T,X,k} = \sum_{m=0}^{\infty}\|\Phi_g^m(g)-\Phi_g^m(0)\|_{T,X,k} \tag{2.30a}$$

$$\leq \sum_{m=0}^{\infty}|\alpha|^m\frac{T^m}{m!}\|g\|_{X,k+m} \tag{2.30b}$$

$$\leq \sum_{m=0}^{\infty}\sum_{j=0}^{\infty}|\alpha|^m T^m X^j |a_{k+m+j}|\frac{(k+m+j)!}{m!j!} \tag{2.30c}$$

[1] Or equivalently, that g has a holomorphic extension on a disk of the complex plane with center 0 and radius R; see Remark 2.4.2.

$$= \sum_{s=0}^{\infty} \sum_{m=0}^{s} |\alpha|^m T^m X^{s-m} |a_{k+s}| \frac{(k+s)!}{m!(s-m)!} \tag{2.30d}$$

$$= \sum_{s=0}^{\infty} (|\alpha| T + X)^s |a_{k+s}| \frac{(k+s)!}{s!}, \tag{2.30e}$$

where (2.30b) is justified by the Contraction Inequalities (Lemma 2.4.1), (2.30c) by equation (2.29), (2.30d) by rearranging the sum and adding over diagonals $s = m + j$ and (2.30e) by taking the binomial expansion of $(|\alpha|T + X)^s$.

By the root test, the above series is convergent, since

$$\limsup_{s\to\infty} \sqrt[s]{(|\alpha|T + X)^s |a_{k+s}| \frac{(k+s)!}{s!}} = (|\alpha|T + X) \cdot \limsup_{n\to\infty} \sqrt[n]{|a_n|} \cdot 1 < \frac{R}{R} = 1.$$

Since the series is convergent, it follows that, for $i < j$,

$$\|u_j - u_i\|_{T,X,k} \leq \sum_{m=i}^{j-1} \|u_{m+1} - u_m\|_{T,X,k} \leq \sum_{m=i}^{\infty} \|u_{m+1} - u_m\|_{T,X,k} \underset{i\to\infty}{\longrightarrow} 0.$$

Hence (u_m) is a Cauchy sequence with respect to the pseudonorm $\|\cdot\|_{T,X,k}$. Since this holds for all $k \in \mathbb{N}$ and $C([0,T], C^\infty(-X, X))$ is complete, it follows that (u_m) has a limit in $\mathbb{X}$. Now, using continuity of $\Phi_g \restriction_{\mathbb{X}}$, we conclude that this limit must be a fixed point of $\Phi_g \restriction_{\mathbb{X}}$. □

Remark 2.4.2. The reader should distinguish between the following two concepts:

- the existence of a holomorphic function, defined in a disk of the complex plane $\mathbb{C}$, which coincides with g at the real axis $\{y = 0\}$;
- the convergence of the construction $u_m = \Phi_g^m(0)$ to a fixed point u, defined in a rectangle of $\mathbb{T} \times \mathbb{R}$, which coincides with g at initial time $\{t = 0\}$.

As seen from Theorem 7, the existence of a holomorphic extension implies convergence to a fixed point. Both these functions (the holomorphic extension and the fixed point) can be depicted by planar diagrams, and both can be seen as extensions of g (see Figure 2.5). However, these functions, and the domains in which they live, are substantially different.

As an immediate corollary of Theorem 7, we have:

Theorem 8 (First global Fréchet space convergence theorem). *Consider the Fréchet space $C(\mathbb{T}, C^\infty(\mathbb{R}))$ with pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Take an initial input $u_0 \in C(\mathbb{T}, \mathcal{X})$ and initial condition $g \in C^\infty(\mathbb{R})$. Assume also that $u_0 = 0$ and g is entire (i.e. has a holomorphic extension to the complex plane). Let $\Phi_g : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by (2.24). Then the sequence (u_m) given by $u_m = \Phi_g^m(u_0)$ converges to a fixed point of Φ_g.*

Proof. Since g is entire, it is analytic at 0 with any radius of convergence R. Thus, by Theorem 7, the sequence u_m converges to a fixed point on any compact rectangle $[0,T]\times[-X,X]$. Therefore, we have convergence of u_m for any pseudonorm $\|\cdot\|_{T,X,k}$, so that we have convergence in $C(\mathbb{T}, C^\infty(\mathbb{R}))$. □

The next step is to generalize Theorem 8 to a larger class of initial functions u_0 (other that $u_0 = 0$). We do that proof in two steps: assume $g = 0$ to establish sufficient conditions on u_0; then consider the more general case $g \in C^\infty(\mathbb{R})$.

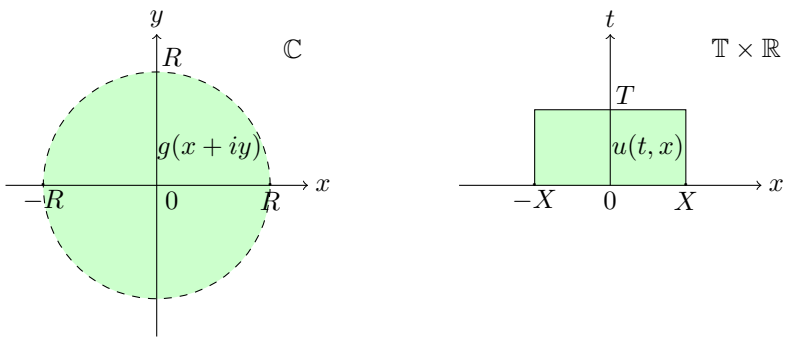


Figure 2.5: On the left: a function $g(x+iy)$ of type $\mathbb{C} \rightharpoonup \mathbb{C}$, defined in a disk, that coincides with g at $\{y = 0, -R < x < R\}$. On the right: a fixed point $u(t,x)$ of type $\mathbb{T} \times \mathbb{R} \rightharpoonup \mathbb{R}$, defined in a rectangle, that coincides with g at $\{t = 0, -X \leq x \leq X\}$. The rectangle and disk dimensions follow the relation $|\alpha|T + X < R$.

Definition 2.4.3 (**Uniform entirety**). We say that a function $u \in C(\mathbb{T}, C^\infty(\mathbb{R}))$ is *uniformly entire* if $u(t,x) = \sum_{j=0}^{\infty} a_j(t)x^j$ for some sequence of functions $(a_j) \in C(\mathbb{R})$ such that

$$\lim \left(\sup_{0 \leq t \leq T} |a_j(t)| \right)^{1/j} = 0 \text{ for all } T \in \mathbb{T}.$$

The motivation for the above terminology is that, for such a function u, the section $x \mapsto u(t,x)$ is entire for all t, and the convergence $\sqrt[j]{|a_j(t)|} \to 0$ is uniform in t.

Theorem 9 (**Global Fréchet space convergence to zero**). *Let $C(\mathbb{T}, C^\infty(\mathbb{R}))$ be the Fréchet space with the family of pseudonorms $\| \cdot \|_{T,X,k}$ given by (2.25). Let also $u_0 \in C(\mathbb{T}, C^\infty(\mathbb{R}))$ be an initial input, and $g \in C^\infty(\mathbb{R})$ be an initial condition. We assume in addition that u_0 is uniformly entire and $g = 0$. Let $\Phi_0 : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by*

$$\Phi_0(u)(t,x) = \alpha \int_0^t \partial_x u(s,x) ds. \tag{2.31}$$

Then the sequence (u_m) given by $u_m = \Phi_0^m(u_0)$ converges to 0.

Proof. To facilitate the exposition we introduce the pseudonorms on a_j given by

$$\|a_j\|_T = \sup_{0 \leq t \leq T} |a_j(t)|, \text{ for } T \in \mathbb{T}. \tag{2.32}$$

We show that $\sum_m \|u_m\|_{T,X,k}$ is a convergent series for any pseudonorm $\| \cdot \|_{T,X,k}$. We have that (see proof of Theorem 7)

$$\sum_{m=0}^{\infty} \|u_m\|_{T,X,k} = \sum_{m=0}^{\infty} \|\Phi_0^m(u_0) - \Phi_0^m(0)\|_{T,X,k} \tag{2.33a}$$

$$\leq \sum_{m=0}^{\infty} \frac{|\alpha|^m T^m}{m!} \|u_0\|_{T,X,k+m} \tag{2.33b}$$

$$= \sum_{m=0}^{\infty} \frac{|\alpha|^m T^m}{m!} \sup_{\substack{0\leq t\leq T \\ |x|\leq X}} \left| \sum_{j=0}^{\infty} \frac{(j+k+m)!}{j!} a_{j+k+m}(t) x^j \right| \tag{2.33c}$$

$$\leq \sum_{m=0}^{\infty} \sum_{j=0}^{\infty} |\alpha|^m T^m X^j \frac{(j+k+m)!}{m!j!} \|a_{j+k+m}\|_T \tag{2.33d}$$

$$= \sum_{s=0}^{\infty} \sum_{m=0}^{s} |\alpha|^m T^m X^{s-m} \frac{(k+s)!}{m!(s-m)!} \|a_{k+s}\|_T \tag{2.33e}$$

$$= \sum_{s=0}^{\infty} (|\alpha|T + X)^s \frac{(k+s)!}{s!} \|a_{k+s}\|_T. \tag{2.33f}$$

By the root test, the above series is convergent, since

$$\left((|\alpha|T + X)^s \frac{(k+s)!}{s!} \right)^{1/s} \underset{s\to\infty}{\longrightarrow} |\alpha|T + X$$

and $\sqrt[s]{\|a_{k+s}\|_T} \underset{s\to\infty}{\longrightarrow} 0$ by assumption. Since $\sum \|u_m\|_{T,X,k}$ is convergent, we then have that $\|u_m\|_{T,X,k} \underset{m\to\infty}{\longrightarrow} 0$ and thus u_m converges to 0, as we wanted to prove. □

We now combine Theorems 8 and 9 to prove our most general result.

Theorem 10 (**Second global Fréchet space convergence theorem**). *Consider the Fréchet space $C(\mathbb{T}, C^\infty(\mathbb{R}))$ with pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Let $u_0 \in C(\mathbb{T}, C^\infty(\mathbb{R}))$ be an initial input and $g \in C^\infty(\mathbb{R})$ be an initial condition. Assume also that u_0 is uniformly entire and that g is entire. Let $\Phi_g : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by (2.24). Then the sequence (u_m) given by $u_m = \Phi_g^m(u_0)$ converges to a fixed point of Φ_g.*

Proof. Let $\Phi_g, \Phi_0 : C(\mathbb{T}, C^\infty(\mathbb{R})) \to C(\mathbb{T}, C^\infty(\mathbb{R}))$ be given by (2.24), (2.31). We observe that, for any $u, v \in C^{0,\infty}(\mathbb{X})$ we have

$$\Phi_g(u+v) = g + \alpha \int_0^t (u+v)_x ds = g + \alpha \int_0^t u_x ds + \alpha \int_0^t v_x ds = \Phi_g(u) + \Phi_0(v).$$

We can then infer that $u_1 = \Phi_g(u_0) = \Phi_g(0 + u_0) = \Phi_g(0) + \Phi_0(u_0)$. Also, $u_2 = \Phi_g(u_1) = \Phi_g(\Phi_g(0) + \Phi_0(u_0)) = \Phi_g^2(0) + \Phi_0^2(u_0)$, and, in general,

$$u_m = \Phi_g^m(0) + \Phi_0^m(u_0).$$

By Theorem 8, $(\Phi_g^m(0))$ converges to a fixed point of Φ_g. By Theorem 9, $(\Phi_0^m(u_0))$ converges to 0. Therefore, (u_m) and $(\Phi_g^m(0))$ have the same limit. In particular, (u_m) converges to a fixed point of Φ_g. □

A nice consequence of the proof is that it allows us to also establish uniqueness in a certain class of functions.

Corollary 2.4.4 (**Uniqueness of uniformly entire fixed points**). *Consider the Fréchet space $C(\mathbb{T}, \mathcal{X})$ with pseudonorms $\|\cdot\|_{T,X,k}$ given by (2.25). Take an initial condition $g \in C^\infty(\mathbb{R})$ and assume also that g is entire. Let $\Phi_g : C(\mathbb{T}, \mathcal{X}) \to C(\mathbb{T}, \mathcal{X})$ be given by (2.24). Then there is at most one uniformly entire fixed point of Φ_g.*

Proof. Let u be any uniformly entire fixed point of Φ_g. By the proof of Theorem 10, we know that $u = \Phi_g^m(u) = \Phi_g^m(0) + \Phi_0^m(u)$. Since $(\Phi_0^m(u))$ converges to 0, we get that $(\Phi_g^m(0))$ converges to u. Thus any uniformly entire fixed point of Φ_g must coincide with the limit of $(\Phi_g^m(0))$. □

2.5 Fourier transform

In this section we shall introduce the Fourier transform, which will lead to a different fixed point approach for obtaining specifications of the analog system. For the moment we focus on presenting a rigorous definition of the Fourier transform, including its domain and co-domain.

As usually, we work on a space $\mathcal{X}$ of functions defined over some spatial domain Ω. We have to consider the cases where Ω is bounded or unbounded separately. For the unbounded case, we shall only treat $\Omega = \mathbb{R}$, whereas for the bounded case we shall treat $\Omega = [0, 2\pi]$, that is, we present the analysis for only one spatial dimension. The higher dimension cases like $\Omega = \mathbb{R}^n$ or $\Omega = [0, 2\pi]^n$ will be omitted for the sake of brevity, but most of the results generalize with additional technical effort.

Unbounded domain: Consider the case where $\mathcal{X}$ is a certain class of functions of type $\mathbb{R} \to \mathbb{C}$. For each such function $f = f(x)$, the Fourier transform, when defined, will be given by the equation

$$(\mathcal{F}f)(\xi) = \hat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ix\xi} f(x) dx. \tag{2.34}$$

Observe that we typically represent f as a function of argument x and $\hat{f}$ as a function of argument ξ; other choices for the Fourier transform are possible, and we include the multiplier of $\frac{1}{\sqrt{2\pi}}$ to get a similar expression for the inverse Fourier transform, which when defined will be given by

$$\left(\mathcal{F}^{-1}\hat{f}\right)(x) = f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{ix\xi} \hat{f}(\xi) d\xi. \tag{2.35}$$

Clearly, equations (2.34) and (2.35) will be valid whenever the integrals converge. For this reason it is typical to consider a condition of absolute integrability for the function f.

Definition 2.5.1 (**Absolute integrability**)**.** A function $f : \mathbb{R} \to \mathbb{C}$ is said to be *absolutely integrable* if $\int_{\mathbb{R}} |f(x)| dx < \infty$. We denote by $L^1(\mathbb{R})$ the space of absolutely integrable functions, quotiented with the equivalence relation of a.e. equality.

Definition 2.5.2 (**Fourier transform in** $L^1(\mathbb{R})$)**.** We denote by $\mathcal{D}_{\mathcal{F}}(\mathbb{R}) \subset L^1(\mathbb{R})$ the space of absolutely integrable functions f such that equation (2.34) also defines an absolutely integrable function $\hat{f}$. We define the Fourier transform operator $\mathcal{F} : \mathcal{D}_{\mathcal{F}}(\mathbb{R}) \to \mathcal{D}_{\mathcal{F}}(\mathbb{R})$ by $\mathcal{F}f = \hat{f}$, given by equation (2.34). We define the inverse Fourier transform $\mathcal{F}^{-1} : \mathcal{D}_{\mathcal{F}}(\mathbb{R}) \to \mathcal{D}_{\mathcal{F}}(\mathbb{R})$ by $\mathcal{F}^{-1}\hat{f} = f$, given by equation (2.35). The fact that $\mathcal{F}$ and $\mathcal{F}^{-1}$ are well-defined bijections in $\mathcal{D}_{\mathcal{F}}(\mathbb{R})$ follows from the Fourier inversion theorem, [Fol95].

Definition 2.5.3 (**Fourier transform in** $\mathcal{S}(\mathbb{R})$)**.** We recall the Schwarz space, denoted by $\mathcal{S}(\mathbb{R})$, which was presented previously in Example 2.2.10 as a Fréchet space. This space consists of infinitely differentiable functions whose derivatives decay rapidly (i.e. faster than any power of x). We then define the Fourier transform and inverse Fourier transform operators $\mathcal{F}, \mathcal{F}^{-1} : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ in a similar fashion.

Remark 2.5.4 (**Fourier transform in** $L^2(\mathbb{R})$)**.** We should mention that the Fourier transform and its inverse can also be defined in the space $L^2(\mathbb{R})$ of *square-integrable functions*, by extending continuously the Fourier transform in $\mathcal{S}(\mathbb{R})$ and using the Plancherel theorem; see for example [Fol95] for details.

Bounded domain: Consider the case where $\mathcal{X}$ is a class of periodic functions of type $[0, 2\pi] \to \mathbb{C}$. For each such function $f = f(x)$, the complex Fourier coefficients, when defined, will be given by the equations

$$(\mathcal{F}f)_k = \hat{f}_k = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{-ikx} f(x) dx, \text{ for } k \in \mathbb{Z}. \tag{2.36}$$

Observe that we represent f as a function of real argument x and $\hat{f}$ as a sequence indexed by integer k; other choices for the Fourier coefficients are possible; when defined, the inverse operation of (2.36) will be given by the Fourier series

$$\left(\mathcal{F}^{-1}\hat{f}\right)(x) = f(x) = \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} \hat{f}_k e^{ikx}. \tag{2.37}$$

Observe that we use the terms 'Fourier transform' and 'inverse Fourier transform' for the unbounded domain and the corresponding terms 'Fourier coefficents' and 'Fourier series' for the bounded domain.

Equations (2.36) and (2.37) will be valid whenever the integral or series converge. As in the unbounded case, it is typical to consider a condition of absolute integrability for f, or absolute summability for $\hat{f}$.

Definition 2.5.5. A function $f : [0, 2\pi] \to \mathbb{C}$ is said to be absolutely square-integrable whenever $\int_0^{2\pi} |f(x)|^2 dx < \infty$. We denote by $L^2([0, 2\pi])$ the space of absolutely square-integrable functions, quotiented with the equivalence relation of a.e. equality.

Definition 2.5.6. A sequence $\hat{f} \in \mathbb{C}^{\mathbb{Z}}$ is said to be absolutely square-summable if $\sum_{k\in\mathbb{Z}} |\hat{f}_k|^2 < \infty$. We denote by $\ell^2(\mathbb{Z})$ the space of absolutely square-summable sequences.

Definition 2.5.7. We define the Fourier transform operator $\mathcal{F} : L^2([0, 2\pi]) \to \ell^2(\mathbb{Z})$, given by equation (2.36) and the inverse Fourier transform operator $\mathcal{F}^{-1} : \ell^2(\mathbb{Z}) \to L^2([0, 2\pi])$, given by equation (2.37).

2.6 Existence, uniqueness and convergence in the Schwarz space

In this section, as a first case study, we consider the situation in which $\mathcal{X} = \mathcal{S}(\mathbb{R})$. The two main advantages of choosing such space are that all operators involved are totally defined and continuous; and we can easily state the fixed point problem in the Fourier space.

Proposition 2.6.1. *Let p be a polynomial in one variable and let $L = p(\partial_x)$ be a linear differential operator acting on $\mathcal{S}(\mathbb{R})$. Then $L : \mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$ is total and continuous. Moreover, if $u \in \mathcal{S}(\mathbb{R})$, then*

$$\mathcal{F}(Lu)(\xi) = p(i\xi)\hat{u}(\xi). \tag{2.38}$$

Proposition 2.6.2. *Let p be a polynomial in one variable and let $L = p(\partial_x)$ be a linear differential operator acting on $\mathcal{S}(\mathbb{R})$. For $g \in \mathcal{S}(\mathbb{R})$, consider the operator $\Phi_g : C(\mathbb{T}, \mathcal{S}(\mathbb{R})) \to C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ given by*

$$\Phi_g(u)(t) = g + \int_0^t Lu(s)ds. \tag{2.39}$$

Then we can define an operator $\hat{\Phi}_{\hat{g}} : C(\mathbb{T}, \mathcal{S}(\mathbb{R})) \to C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$*, given by*

$$\hat{\Phi}_{\hat{g}}(\hat{u})(t) = \hat{g} + \int_0^t p(i\xi)\hat{u}(s)ds. \tag{2.40}$$

Moreover, for $u \in C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$*,* u *is a fixed point of* Φ_g *if and only if* $\hat{u}$ *is a fixed point of* $\hat{\Phi}_{\hat{g}}$*.*

Therefore, we shall try to obtain existence and uniqueness results for fixed points of $\hat{\Phi}_{\hat{g}}$, defined in (2.40). As usual, we begin by proving contraction inequalities. We observe that $\mathcal{S}(\mathbb{R})$ is a Fréchet space, and we shall use a slighty different (yet equivalent) choice of pseudonorms from Example 2.2.10, namely for $a, b \in \mathbb{N}$ we consider the pseudonorm

$$\|u\|_{a,b} = \sup_{x\in\mathbb{R}} \left|(1+|x|)^a \partial_x^b u(x)\right|, \tag{2.41}$$

so that $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ is a Fréchet space with pseudonorms

$$\|u\|_{a,b,T} = \sup_{0\le t\le T} \|u(t)\|_{a,b}. \tag{2.42}$$

Proposition 2.6.3. *Let* p *be a polynomial of degree* m*. There is a constant* C*, depending on* p *only, such that for any* $\xi \in \mathbb{R}$*,* $k \in \mathbb{N}$ *and* $j \in \mathbb{N}$*, we have*

$$\left|\partial_\xi^j (p(i\xi)^k)\right| \le (mk)^j C^k (1+|\xi|)^{mk}. \tag{2.43}$$

Proposition 2.6.4. *Let* p *be a polynomial of degree* m *and* $g \in \mathcal{S}(\mathbb{R})$*. There is a constant* C*, depending on* p *only, such that for any pseudonorm* $\|\cdot\|_{a,b,T}$ *and any* $k \in \mathbb{N}$*,* $\hat{u}$*,* $\hat{v} \in C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$*, we have*

$$\left\|\hat{\Phi}_{\hat{g}}^k(\hat{u}) - \hat{\Phi}_{\hat{g}}^k(\hat{v})\right\|_{a,b,T} \le \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} \|\hat{u}-\hat{v}\|_{a+km,j,T}. \tag{2.44}$$

Proof. First observe that $p(i\xi)^k(\hat{u}-\hat{v})$ is infinitely differentiable in ξ, with

$$\partial_\xi^b \left(p(i\xi)^k(\hat{u}-\hat{v})\right) = \sum_{j=0}^{b} \binom{b}{j} \partial_\xi^j \left(p(i\xi)^k\right) \partial_\xi^{b-j}(\hat{u}-\hat{v}). \tag{2.45}$$

We have

$$\left\|\hat{\Phi}_{\hat{g}}^k(\hat{u}) - \hat{\Phi}_{\hat{g}}^k(\hat{v})\right\|_{a,b,T} = \left\|\int_0^t \cdots \int_0^{s_{k-1}} p(i\xi)^k(\hat{u}-\hat{v}) ds_k \dots ds_1\right\|_{a,b,T} \tag{2.46a}$$

$$\le \sup_{0\le t\le T} \sup_{\xi\in\mathbb{R}} \int_0^t \cdots \int_0^{s_{k-1}} (1+|\xi|)^a \left|\partial_\xi^b \left(p(i\xi)^k(\hat{u}-\hat{v})\right)\right| ds_k \dots ds_1 \tag{2.46b}$$

$$\le \frac{T^k}{k!} \sup_{0\le t\le T} \sup_{\xi\in\mathbb{R}} (1+|\xi|)^a \sum_{j=0}^{b} \binom{b}{j} \left|\partial_\xi^j \left(p(i\xi)^k\right) \partial_\xi^{b-j}(\hat{u}-\hat{v})\right| \tag{2.46c}$$

$$\leq \frac{T^k}{k!} \sup_{0\leq t\leq T} \sup_{\xi\in\mathbb{R}} \sum_{j=0}^{b} \binom{b}{j} (mk)^j C^k (1+|\xi|)^{a+mk} \left|\partial_\xi^{b-j}(\hat{u}-\hat{v})\right| \tag{2.46d}$$

$$\leq \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} \|\hat{u}-\hat{v}\|_{a+km,j,T}, \tag{2.46e}$$

where (2.46c) is justified by equation (2.45) and (2.46d) by Proposition 2.6.3. This concludes the proof. □

Remark 2.6.5. Once more, we arrive at a situation similar to what occured in the previous sections: our contraction inequalities relate pseudonorms $\|\cdot\|_{a,b,T}$ with pseudonorms $\|\cdot\|_{a',b',T}$, for some a' higher than a (in this case, $a' = a+km$) and some b' smaller than b (in this case, all $b' = 0,\dots,b$). As we have seen before, a possible next step would be to impose conditions on $\hat{u}$ and $\hat{v}$ that ensure

$$\left\|\hat{\Phi}_{\hat{g}}^k(\hat{u}) - \hat{\Phi}_{\hat{g}}^k(\hat{v})\right\|_{a,b,T} \underset{k\to\infty}{\longrightarrow} 0.$$

Proposition 2.6.6. *Let p be a polynomial of degree m and $g \in \mathcal{S}(\mathbb{R})$. Let $\hat{u}$, $\hat{v} \in C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ with the following property: for all $b \in \mathbb{N}$, $T \in \mathbb{T}$, there is a constant $C_{b,T}$ such that, for all $a \in \mathbb{N}$,*

$$\|\hat{u}\|_{a,b,T}, \|\hat{v}\|_{a,b,T} \leq C_{b,T}^a.$$

Then

$$\left\|\hat{\Phi}_{\hat{g}}^k(\hat{u}) - \hat{\Phi}_{\hat{g}}^k(\hat{v})\right\|_{a,b,T} \underset{k\to\infty}{\longrightarrow} 0.$$

Proof. We have

$$\left\|\hat{\Phi}_{\hat{g}}^k(\hat{u}) - \hat{\Phi}_{\hat{g}}^k(\hat{v})\right\|_{a,b,T} \leq \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} \|\hat{u}-\hat{v}\|_{a+km,j,T} \tag{2.47a}$$

$$\leq \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} \|\hat{u}\|_{a+km,j,T} + \|\hat{v}\|_{a+km,j,T} \tag{2.47b}$$

$$\leq \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} C_{j,T}^{a+km} + C_{j,T}^{a+km} \tag{2.47c}$$

$$\leq \frac{(2mk)^b (CT)^k}{k!} 2 \left(\sum_{j=0}^{b} C_{j,T}\right)^{a+km} \underset{k\to\infty}{\longrightarrow} 0, \tag{2.47d}$$

where (2.47a) is justified by Proposition 2.6.4, (2.47b) by triangle inequality and (2.47c) by the growth bounds on the pseudonorms of $\hat{u}$ and $\hat{v}$. This concludes the proof. □

Theorem 11 (Fixed points of $\hat{\Phi}_{\hat{g}}$: existence and convergence in $\mathcal{S}(\mathbb{R})$). *Let p be a polynomial of degree m and $g \in \mathcal{S}(\mathbb{R})$. Suppose that for all $b \in \mathbb{N}$, there is a constant C_b such that, for all $a \in \mathbb{N}$,*

$$\|\hat{g}\|_{a,b} \leq C_b^a.$$

Then the sequence $\hat{u}_k = \hat{\Phi}_{\hat{g}}^k(0)$ converges in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ to a fixed point $\hat{u}^*$ of $\hat{\Phi}_{\hat{g}}$.

Proof. Fix a, b, T. We observe that

$$\sum_{k=0}^{\infty} \|\hat{u}_{k+1} - \hat{u}_k\|_{a,b,T} = \sum_{k=0}^{\infty} \|\hat{\Phi}_{\hat{g}}^k(\hat{g}) - \hat{\Phi}_{\hat{g}}^k(0)\|_{a,b,T} \tag{2.48a}$$

$$\leq \sum_{k=0}^{\infty} \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} \|\hat{g}\|_{a+km,j,T} \tag{2.48b}$$

$$\leq \sum_{k=0}^{\infty} \frac{(2mk)^b (CT)^k}{k!} \sum_{j=0}^{b} C_j^{a+km} \tag{2.48c}$$

$$\leq \sum_{k=0}^{\infty} \frac{(2mk)^b (CT)^k}{k!} \left(\sum_{j=0}^{b} C_j \right)^{a+km} < \infty, \tag{2.48d}$$

where (2.48b) is justified by Proposition 2.6.4 and (2.48c) by the growth bounds on the pseudonorms of $\hat{g}$. The summability of the last series can be justified, for example, by the ratio test (or noting that the $k!$ factor greatly dominates the remaining factors).

Repeating the reasoning of previous proofs (cf. Theorem 7), we then conclude that, for $i < j$, $\|\hat{u}_j - \hat{u}_i\|_{a,b,T} \underset{i\to\infty}{\longrightarrow} 0$, so that $(\hat{u}_k)$ is a Cauchy sequence and thus it must converge to some limit $\hat{u}^*$, which must be a fixed point of $\hat{\Phi}_{\hat{g}}$. □

Remark 2.6.7. For a given $\hat{g} \in \mathcal{S}(\mathbb{R})$, let us consider the sequence u_k where $\hat{u}_0 = 0$ and $\hat{u}_{k+1} = \hat{\Phi}_{\hat{g}}(\hat{u}_k)$. It is easy to check that $u_1(t,\xi) = \hat{g}(\xi)$, $u_2(t,\xi) = \hat{g}(\xi) + p(i\xi)t\hat{g}(\xi)$ and more generally

$$\hat{u}_k = \sum_{i=0}^{k-1} \frac{p(i\xi)^k t^k}{k!} \hat{g}(\xi);$$

this can be proven by induction on k, or simply by applying the Fourier Transform to the result on Lemma 2.3.1. Thus, we have convergence to a fixed point if and only if the series

$$\sum_{k=0}^{\infty} \frac{p(i\xi)^k t^k}{k!} \hat{g}(\xi) \tag{2.49}$$

is convergent in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$.

It is a relevant question whether the growth conditions on g in Theorem 11 can be relaxed in some manner. We present an example where the series in (2.49) is not absolutely convergent, and thus it suggests a negative answer.

Example 2.6.8. Let p be a polynomial of degree $m > 0$, which can be written as $p(i\xi) = C_m\xi^m + \ldots + C_1\xi + C_0$, where C_i are complex numbers and the leading coefficient $C_m \neq 0$. We can make an estimate and assume there exist positive constants C and X such that

$$|p(i\xi)| \geq C\xi^m, \quad \text{for } \xi \geq X. \tag{2.50}$$

Now consider $\hat{g} \in \mathcal{S}(\mathbb{R})$ such that

$$\hat{g}(\xi) = e^{-\xi}, \quad \text{for } \xi \geq X; \tag{2.51}$$

since the decay is exponential for $\xi \to +\infty$, we can construct such a $\hat{g}$. Next we consider k large enough, namely $k \geq X$, and estimate one particular pseudonorm ($a = 0$, $b = 0$, some T),

$$\left\| \frac{p(i\xi)^k t^k}{k!} \hat{g} \right\|_{0,0,T} = \sup_{0 \leq t \leq T} \sup_{\xi \in \mathbb{R}} \left| \frac{p(i\xi)^k t^k}{k!} \hat{g}(\xi) \right| \tag{2.52a}$$

$$\geq \sup_{\xi \geq X} \left| \frac{p(i\xi)^k T^k}{k!} e^{-\xi} \right| \tag{2.52b}$$

$$\geq \sup_{\xi \geq X} \frac{(CT\xi^m)^k}{k!} e^{-\xi} \tag{2.52c}$$

$$\geq \frac{(CTk^m)^k e^{-k}}{k!} \tag{2.52d}$$

$$\geq \frac{(CTk^m)^k e^{-k}}{ek^{k+1/2}e^{-k}} \tag{2.52e}$$

$$= \frac{(CT)^k}{e} k^{(m-1)k-1/2}, \tag{2.52f}$$

where (2.52c) is justified by the bound (2.50) on $p(i\xi)$, (2.52d) by evaluating at $\xi = k$ and (2.52e) by Stirling's approximation [Rud76, Chapter 8],

$$\sqrt{2\pi} k^{k+\frac{1}{2}} e^{-k} \leq k! \leq e k^{k+\frac{1}{2}} e^{-k}. \tag{2.53}$$

If T is taken to be large enough, namely $T > 1/C$, then the series with terms given as $\frac{(CT)^k}{e} k^{(m-1)k-1/2}$ diverges, and therefore (2.49) is not absolutely convergent for $k \to \infty$, that is,

$$\lim_{K \to \infty} \sum_{k=0}^{K} \left\| \frac{p(i\xi)^k t^k}{k!} \hat{g} \right\|_{0,0,T} = +\infty.$$

Of course, this does not tell us definitely whether there is convergence to the fixed point; we only proved that (2.49) is not an absolutely convergent series, but not that it is not convergent (as per Remark 2.6.7). In particular, the Riemann rearrangement theorem suggests that the sum may depend on the order we sum its terms. In any case, it can be seen that our choice of g does not fulfill the conditions on Theorem 11, and in fact for any $b \in \mathbb{N}$, $\|g\|_{a,b}$ grows like $(ae)^{-a}$.

We next consider a slightly different approach and reframe our problem in the larger space of $C^\infty(\mathbb{R})$ functions. Recall that this is a Fréchet space with pseudonorms

$$\|u\|_{M,b} = \sup_{|x| \leq M} \left| \partial_x^b u(x) \right|, \tag{2.54}$$

so that $C(\mathbb{T}, C^\infty(\mathbb{R}))$ is a Fréchet space with pseudonorms

$$\|u\|_{M,b,T} = \sup_{0 \leq t \leq T} \|u(t)\|_{M,b}. \tag{2.55}$$

We observe that $\mathcal{S}(\mathbb{R})$ is a linear subspace of $C(\mathbb{R})$; one can also see that the topology induced by the pseudonorms of $\mathcal{S}(\mathbb{R})$ is finer than the topology induced by the pseudonorms of $C^\infty(\mathbb{R})$. Thus, if $(u_n)_{n \in \mathbb{N}}$ is a sequence in $\mathcal{S}(\mathbb{R})$ converging to $u \in \mathcal{S}(\mathbb{R})$, then we also have that u_n converges to u

in $C^\infty(\mathbb{R})$. However, the converse does not hold, as the following example shows.

Example 2.6.9. Consider the sequence $u_n(x) = e^{-(x-n)^2}$, given by the translation of the Gaussian e^{-x^2} to the right by n units. It is obvious that $u_n \in C^\infty(\mathbb{R})$ for all n; also, the fast decay at infinity ensures that $u_n \in \mathcal{S}(\mathbb{R})$ as well. Moreover, one can see that, since the Gaussian wave 'travels to infinity' as $n \to \infty$, then the sequence u_n and any of its derivatives vanish in every fixed compact interval. In other words, $u_n \to 0$ in the topology induced by the $C^\infty(\mathbb{R})$ pseudonorms.

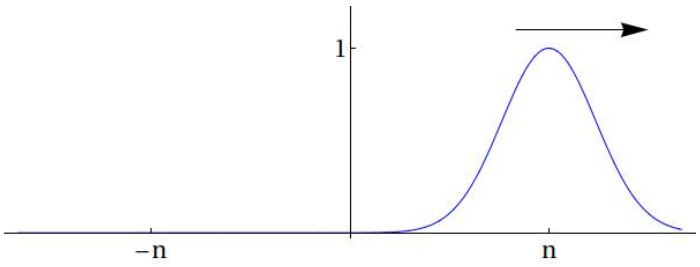


Figure 2.6: Plot of a Gaussian wave $u_n(x)$.

However, such convergence does not hold in the $\mathcal{S}(\mathbb{R})$-topology; in fact, just by looking at the first pseudonorm $\|\cdot\|_{0,0}$ one sees that

$$\|u_n\|_{0,0} = \sup_{x\in\mathbb{R}} |u_n(x)| = u_n(n) = 1 \not\to 0,$$

and thus $u_n \not\to 0$ in $\mathcal{S}(\mathbb{R})$.

The arguments presented above also apply in the corresponding stream spaces $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ and $C(\mathbb{T}, C^\infty(\mathbb{R}))$; moreover, when $\hat{g} \in \mathcal{S}(\mathbb{R})$ (and more generally for $\hat{g} \in C^\infty(\mathbb{R})$), we can extend the operator $\hat{\Phi}_{\hat{g}}$ given by (2.40) and think of it as acting on $C(\mathbb{T}, C^\infty(\mathbb{R}))$.

Proposition 2.6.10. *Let p be a polynomial of degree m and $g \in \mathcal{S}(\mathbb{R})$. There is a constant C, depending on p only, such that for any pseudonorm $\|\cdot\|_{M,b,T}$ and any $k \in \mathbb{N}$, $\hat{u}$, $\hat{v} \in C(\mathbb{T}, C^\infty(\mathbb{R}))$, we have*

$$\left\|\hat{\Phi}^k_{\hat{g}}(\hat{u}) - \hat{\Phi}^k_{\hat{g}}(\hat{v})\right\|_{M,b,T} \leq \frac{(2mk)^b (CT(1+M)^m)^k}{k!} \sum_{j=0}^{b} \|\hat{u} - \hat{v}\|_{M,j,T}. \tag{2.56}$$

Proof. We have

$$\left\|\hat{\Phi}^k_{\hat{g}}(\hat{u}) - \hat{\Phi}^k_{\hat{g}}(\hat{v})\right\|_{M,b,T} = \left\|\int_0^t \cdots \int_0^{s_{k-1}} p^k(i\xi)(\hat{u} - \hat{v}) ds_k \dots ds_1\right\|_{M,b,T} \tag{2.57a}$$

$$\leq \sup_{0\leq t\leq T} \sup_{|\xi|\leq M} \int_0^t \cdots \int_0^{s_{k-1}} \left|\partial^b_\xi \left(p^k(i\xi)(\hat{u} - \hat{v})\right)\right| ds_k \dots ds_1 \tag{2.57b}$$

$$\leq \frac{T^k}{k!} \sup_{0\leq t\leq T} \sup_{|\xi|\leq M} \sum_{j=0}^{b} \binom{b}{j} \left|\partial^j_\xi \left(p^k(i\xi)\right) \partial^{b-j}_\xi (\hat{u} - \hat{v})\right| \tag{2.57c}$$

$$\leq \frac{T^k}{k!} \sup_{0\leq t\leq T} \sup_{|\xi|\leq M} \sum_{j=0}^{b} \binom{b}{j} (mk)^j C^k (1 + |\xi|)^{mk} \left|\partial^{b-j}_\xi (\hat{u} - \hat{v})\right| \tag{2.57d}$$

$$\leq \frac{(2mk)^b (CT(1+M)^m)^k}{k!} \sum_{j=0}^{b} \|\hat{u} - \hat{v}\|_{M,j,T}, \tag{2.57e}$$

where (2.57c) is justified by Equation (2.45) and (2.57d) by Proposition 2.6.3. This concludes the proof. □

With the previous results in hand, we proceed to prove existence and uniqueness of a fixed point of $\hat{\Phi}_{\hat{g}}$ defined in $C(\mathbb{T}, C^\infty(\mathbb{R}))$.

Theorem 12 (Fixed points of $\hat{\Phi}_{\hat{g}}$: existence, uniqueness and convergence in $C^\infty(\mathbb{R})$). *Let p be a polynomial of degree m and $g \in \mathcal{S}(\mathbb{R})$. Then there is exactly one fixed point of $\hat{\Phi}_{\hat{g}}$ in $C(\mathbb{T}, C^\infty(\mathbb{R}))$, and moreover, for any $\hat{u}_0 \in C(\mathbb{T}, C^\infty(\mathbb{R}))$, we have $\hat{\Phi}^k_{\hat{g}}(\hat{u}_0) \underset{k\to\infty}{\longrightarrow} \hat{u}^*$ in $C(\mathbb{T}, C^\infty(\mathbb{R}))$, with $\hat{u}^*$ denoting the fixed point.*

Proof. We first simplify the bound given by Proposition 2.6.10 by noting that, for any $M, b, T \in \mathbb{N}$, there is a constant $C_{p,M,b,T}$ (depending on p, M, b and T) such that, for all $k \in \mathbb{N}$,

$$\left\| \hat{\Phi}^k_{\hat{g}}(\hat{u}) - \hat{\Phi}^k_{\hat{g}}(\hat{v}) \right\|_{M,b,T} \leq \frac{C^k_{p,M,b,T}}{k!} \sum_{j=0}^{b} \|\hat{u} - \hat{v}\|_{M,j,T}. \tag{2.58}$$

Existence and convergence: Take an arbitrary $\hat{u}_0 \in C(\mathbb{T}, C^\infty(\mathbb{R}))$; we shall show that the sequence $\hat{u}_k := \hat{\Phi}^k_{\hat{g}}(\hat{u}_0)$ has a limit.

Fix $M, b, T \in \mathbb{N}$. As usual, we observe that

$$\sum_{k=0}^{\infty} \|\hat{u}_{k+1} - \hat{u}_k\|_{M,b,T} = \sum_{k=0}^{\infty} \|\hat{\Phi}^k_{\hat{g}}(\hat{u}_1) - \hat{\Phi}^k_{\hat{g}}(\hat{u}_0)\|_{M,b,T} \tag{2.59a}$$

$$\leq \sum_{k=0}^{\infty} \frac{C^k_{p,M,b,T}}{k!} \sum_{j=0}^{b} \|\hat{u}_1 - \hat{u}_0\|_{M,j,T} \tag{2.59b}$$

$$= e^{C_{p,M,b,T}} \sum_{j=0}^{b} \|\hat{u}_1 - \hat{u}_0\|_{M,j,T} < \infty, \tag{2.59c}$$

where (2.59b) is justified by Equation (2.58).

Repeating the reasoning of previous proofs (cf. Theorem 7), we then conclude that $\|\hat{u}_j - \hat{u}_i\|_{a,b,T} \underset{i,j\to\infty}{\longrightarrow} 0$, so that $(\hat{u}_k)$ is a Cauchy sequence and thus it must converge to some limit $\hat{u}^*$, which must be (by continuity) a fixed point of $\hat{\Phi}_{\hat{g}}$.

Uniqueness: Let $\hat{u}^*$, $\hat{v}^* \in C(\mathbb{T}, C^\infty(\mathbb{R}))$ be fixed points of $\hat{\Phi}_{\hat{g}}$. We shall prove that they coincide by showing that, for every M, b, T, we have $\|\hat{u}^* - \hat{v}^*\|_{M,b,T} = 0$. This will be done by induction on $b \in \mathbb{N}$.

Base step: Let $b = 0$ and let k be large enough such that $\frac{C^k_{p,M,0,T}}{k!} < 1$. Then

$$\|\hat{u}^* - \hat{v}^*\|_{M,0,T} = \|\hat{\Phi}^k_{\hat{g}}(\hat{u}^*) - \hat{\Phi}^k_{\hat{g}}(\hat{v}^*)\|_{M,0,T} \leq \frac{C^k_{p,M,0,T}}{k!} \|\hat{u}^* - \hat{v}^*\|_{M,0,T}, \tag{2.60}$$

which implies $\|\hat{u}^* - \hat{v}^*\|_{M,0,T} = 0$.

Induction step: Assume $\|\hat{u}^* - \hat{v}^*\|_{M,j,T} = 0$ for $j = 0, \dots, b$. Let k be large enough such that $\frac{C^k_{p,M,b+1,T}}{k!} < 1$. Then

$$\|\hat{u}^* - \hat{v}^*\|_{M,b+1,T} = \|\hat{\Phi}^k_{\hat{g}}(\hat{u}^*) - \hat{\Phi}^k_{\hat{g}}(\hat{v}^*)\|_{M,b+1,T} \leq \frac{C^k_{p,M,b+1,T}}{k!} \sum_{j=0}^{b+1} \|\hat{u}^* - \hat{v}^*\|_{M,j,T}$$
$$= \frac{C^k_{p,M,b+1,T}}{k!} \|\hat{u}^* - \hat{v}^*\|_{M,b+1,T},$$

where the last step is justified by induction hypothesis; the inequality then implies $\|\hat{u}^* - \hat{v}^*\|_{M,b+1,T} = 0$.

Since we have $\|\hat{u}^* - \hat{v}^*\|_{M,b,T} = 0$, we can conclude by point separability that $\hat{u}^* = \hat{v}^*$, so that we have uniqueness. □

From Theorem 12, we get existence, uniqueness and convergence to a fixed point in the space $C(\mathbb{T}, C^\infty(\mathbb{R}))$. Therefore, one concludes that there is at most one fixed point of $\hat{\Phi}_{\hat{g}}$ in the subspace $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$. The next logical step is to question whether the fixed point in $C(\mathbb{T}, C^\infty(\mathbb{R}))$, obtained through Theorem 12, also belongs in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$, and if so, whether the convergence to that fixed point is preserved in the finer topology.

As an attempt to answer this question, we first look at the linear (in $\hat{u}$) evolution problem

$$\hat{u}_t(t, \xi) = p(i\xi)\hat{u}(t, \xi), \quad \hat{u}(0, \xi) = \hat{g}(\xi);$$

if we treat this as a decoupled system, then we can use the classical expression

$$\hat{u}^*(t, \xi) = e^{p(i\xi)t}\hat{g}(\xi) \tag{2.61}$$

as the solution to our problem. We can easily check that $\hat{u}^*$ must be the fixed point of $\hat{\Phi}_{\hat{g}}$ in the space $C(\mathbb{T}, C^\infty(\mathbb{R}))$. Next, we determine whether $\hat{u}^*$ is in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$; as we will see, this can be accomplished by enforcing a condition on the polynomial p. Recall that, for a complex number z, $\mathrm{Re}(z)$ denotes the real part of z.

Proposition 2.6.11. *Let p be a polynomial of degree m and $g \in \mathcal{S}(\mathbb{R})$. Suppose that there is a constant C such that, for all $\xi \in \mathbb{R}$, $\mathrm{Re}(p(i\xi)) \leq C$. Then $\hat{u}^*$ defined as in (2.61) is in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$.*

Proof. First observe that $\hat{u}^*$ is infinitely differentiable in ξ, with

$$\partial^b_\xi \left(e^{p(i\xi)t}\hat{g} \right) = \sum_{k=0}^{b} \binom{b}{k} \partial^k_\xi \left(e^{p(i\xi)t} \right) \partial^{b-k}_\xi \hat{g}; \tag{2.62}$$

moreover, the term $\partial^k_\xi \left(e^{p(i\xi)t} \right)$ can be seen to be of the form $q_k(t, \xi) \cdot e^{p(i\xi)t}$, for some q_k polynomial in t and ξ. Let m_k be the degree[2] of q_k. Now we can show that all pseudonorms are finite,

$$\|\hat{u}^*\|_{a,b,T} = \sup_{0 \leq t \leq T} \sup_{\xi \in \mathbb{R}} \left| (1 + |\xi|)^a \partial^b_\xi \left(e^{p(i\xi)t}\hat{g} \right) \right| \tag{2.63a}$$
$$\leq \sup_{0 \leq t \leq T} \sup_{\xi \in \mathbb{R}} \sum_{k=0}^{b} \left| \binom{b}{k} (1 + |\xi|)^a q_k(t, \xi) e^{p(i\xi)t} \partial^{b-k}_\xi \hat{g} \right| \tag{2.63b}$$

[2] it is fairly trivial to prove that $m_k = k(m-1)$, but we do not need this fact.

$$\leq e^{CT} \sum_{k=0}^{b} \sup_{0\leq t\leq T} \sup_{\xi\in\mathbb{R}} \left| \binom{b}{k} (1+|\xi|)^a q_k(t,\xi) \partial_\xi^{b-k} \hat{g} \right| < \infty, \tag{2.63c}$$

where (2.63b) is justified by Equation (2.62); (2.63c) by the hypothesis that $\mathrm{Re}(p(i\xi)) \leq C$ and thus $|e^{p(i\xi)t}| \leq e^{Ct}$ for all t, ξ; and to justify the finiteness of the last term, we can bound each of the terms in the sum by a weighted sum of the pseudonorms $\|\hat{g}\|_{a',b-k}$ for $a \leq a' \leq a + m_k$. We conclude that $\hat{u}^* \in C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$. □

Remark 2.6.12. We can actually see that the condition $\mathrm{Re}(p(i\xi)) \leq C$ is necessary for $e^{p(i\xi)t}\hat{g}$ to be in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$ for all $\hat{g}$. Let p be a polynomial such that $\mathrm{Re}(p(i\xi))$ is positively unbounded; thus we can assume that $\mathrm{Re}(p(i\xi)) \geq C|\xi|^k$ for some constants $C > 0$ and $k \in \mathbb{N}^+$, and for large positive ξ, or large negative ξ; without loss of generality we assume this holds for large positive ξ, say $\xi \geq X$. Now take $\hat{g} \in \mathcal{S}(\mathbb{R})$ such that

$$\hat{g}(\xi) = e^{-C\xi^k}, \quad \text{for } \xi \geq X;$$

since the decay is exponential for $\xi \to +\infty$, we can construct such a $\hat{g}$. Finally we prove that $e^{p(i\xi)t}\hat{g} \notin C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$, just by looking at one of its pseudonorms ($a = 0$, $b = 0$, $T = 2$),

$$\begin{aligned}
\|e^{p(i\xi)t}\hat{g}\|_{0,0,2} &= \sup_{0\leq t\leq 2} \sup_{\xi\in\mathbb{R}} |e^{p(i\xi)t}\hat{g}| \\
&\geq \sup_{0\leq t\leq 2} \sup_{\xi\geq X} |e^{p(i\xi)t} e^{-C\xi^k}| \\
&= \sup_{0\leq t\leq 2} \sup_{\xi\geq X} e^{\mathrm{Re}(p(i\xi))t} e^{-C\xi^k} \\
&\geq \sup_{0\leq t\leq 2} \sup_{\xi\geq X} e^{C(t-1)\xi^k} = +\infty.
\end{aligned}$$

We have seen that, under some conditions on the polynomial $p(i\xi)$, the solution to our problem given by (2.61) does indeed belong in the smaller space $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$. However, it is a much harder question whether the construction presented in Theorem 12 converges to that fixed point in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$. Example 2.6.9 shows that convergence in $C(\mathbb{T}, C^\infty(\mathbb{R}))$ does not entail convergence in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$, and Example 2.6.8 shows that, for some choices of $\hat{g}$, the construction in Theorem 11 produces a series (2.49) which is not absolutely convergent in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$, even though Theorem 12 implies that such series is absolutely convergent in $C(\mathbb{T}, C^\infty(\mathbb{R}))$. Based on these results, we would conjecture that, in general, there is no convergence in the finer topology, even with the bound on the polynomial $p(i\xi)$. However, a proof of such statement, or of its negation, still eludes us.

2.7 Existence, uniqueness and convergence in the h^∞-space

In this section, as a second case study, we consider the case in which $\mathcal{X} = C_p^\infty([0, 2\pi])$. In other words, $\mathcal{X}$ is the space of infinitely differentiable functions g of type $[0, 2\pi] \to \mathbb{C}$ such that

$$g^{(k)}(0) = g^{(k)}(2\pi) \quad \text{for all } k \in \mathbb{N}. \tag{2.64}$$

We refer to (2.64) as *periodic boundary conditions*; the motivation is that such a function g can be extended to an infinitely differentiable function $\tilde{g} \in C^\infty(\mathbb{R})$ such that $\tilde{g}(x) = \tilde{g}(x + 2\pi)$ for all $x \in \mathbb{R}$.

We also remark that $C_p^\infty([0,2\pi])$ is a Fréchet space under the pseudonorms

$$\|g\|_M^2 = \int_0^{2\pi} \left| \frac{d^M}{dx^M} g(x) \right|^2 dx.$$

The reason for taking integral norms (instead of supremum norms) will be made clear shortly. We introduce the subspace $h^\infty(\mathbb{Z})$ of $\ell^2(\mathbb{Z})$ given by those sequences which decay faster than algebraically, in the sense that $\hat{g} \in h^\infty(\mathbb{Z})$ if and only if the quantities

$$\|\hat{g}\|_M^2 = \sum_{k \in \mathbb{Z}} k^{2M} |\hat{g}_k|^2 \tag{2.65}$$

are finite for every $M \in \mathbb{N}$. It should be mentioned that (2.65) define a family of pseudonorms and $h^\infty(\mathbb{Z})$ is a Fréchet space. Moreover, we have the following result relating $C_p^\infty([0,2\pi])$ and $h^\infty(\mathbb{Z})$.

Proposition 2.7.1. *Let $\mathcal{F} : L^2([0,2\pi]) \to \ell^2(\mathbb{Z})$ and $\mathcal{F}^{-1} : \ell^2(\mathbb{Z}) \to L^2([0,2\pi])$ be the Fourier transforms on Definition 2.5.7.*

1. *$\mathcal{F}(C_p^\infty([0,2\pi])) = h^\infty(\mathbb{Z})$ and $\mathcal{F}^{-1}(h^\infty(\mathbb{Z})) = C_p^\infty([0,2\pi])$.*
2. *If $g \in C_p^\infty([0,2\pi])$ and $\hat{g} = \mathcal{F}(g)$, then for all $M \in \mathbb{N}$ we have*

$$\|g\|_M = \|\hat{g}\|_M.$$

In other words, the restrictions of $\mathcal{F}$ to $C_p^\infty([0,2\pi])$ and $\mathcal{F}^{-1}$ to $h^\infty(\mathbb{Z})$ are inverses to respect to each other and are isometries.

Most of this section will consist of proving the analog of some results on Section 2.6 for the bounded domain case. We begin by showing how to state the fixed point problem in the space of Fourier coefficients.

Proposition 2.7.2. *Let p be a polynomial in one variable and let $L = p(\partial_x)$ be a linear differential operator acting on $C_p^\infty([0,2\pi])$. Then $L : C_p^\infty([0,2\pi]) \to C_p^\infty([0,2\pi])$ is total and continuous. Moreover, if $g \in C_p^\infty([0,2\pi])$, then*

$$(\mathcal{F}(Lg))_k = p(ik)\hat{g}_k. \tag{2.66}$$

Proposition 2.7.3. *Let p be a polynomial in one variable and let $L = p(\partial_x)$ be a linear differential operator acting on $C_p^\infty([0,2\pi])$. For $g \in C_p^\infty([0,2\pi])$, consider the operator $\Phi_g : C(\mathbb{T}, C_p^\infty([0,2\pi])) \to C(\mathbb{T}, C_p^\infty([0,2\pi]))$ given by*

$$\Phi_g(u)(t) = g + \int_0^t Lu(s)ds. \tag{2.67}$$

Then we can define an operator $\hat{\Phi}_{\hat{g}} : C(\mathbb{T}, h^\infty(\mathbb{Z})) \to C(\mathbb{T}, h^\infty(\mathbb{Z}))$, given by

$$\hat{\Phi}_{\hat{g}}(\hat{u})(t)_k = \hat{g}_k + \int_0^t p(ik)\hat{u}_k(s)ds. \tag{2.68}$$

Moreover, for $u \in C(\mathbb{T}, C_p^\infty([0,2\pi]))$, u is a fixed point of Φ_g if and only if $\hat{u}$ is a fixed point of $\hat{\Phi}_{\hat{g}}$.

Remark 2.7.4. At this point, we can give an additional motivation for requiring, at the beginning of this section, that all derivatives of functions in $\mathcal{X}$ must coincide at the boundary. On the usual

problems in PDEs, typical boundary conditions involve only a finite number of derivatives at the endpoints. However, we are interested in a framework in which the operator $L = p(\partial_x)$ is total. Thus it makes sense to impose that any function in $\mathcal{X}$ be infinitely differentiable. Moreover, if we only consider a finite number of boundary conditions in the definition of $\mathcal{X}$, it may happen that $Lg \notin \mathcal{X}$ for some $g \in \mathcal{X}$, as the following example suggest.

Example 2.7.5. Let $L = \partial_x^2$, so fixed points of Φ_g correspond to solutions of the heat equation $u_t = u_{xx}$ with initial data g. The typical periodic boundary conditions for the heat equation are that $u(t,0) = u(t,2\pi)$ and $u_x(t,0) = u_x(t,2\pi)$. One could naively start by considering the space $\mathcal{X}$ of functions g of type $[0,2\pi] \to \mathbb{C}$ which are (at least once) continuously differentiable and with a periodic condition for g and g'. If we take initial data $g(x) = x(x-\pi)(x-2\pi)$ which is infinitely differentiable, one can see that $g(0) = g(2\pi) = 0$ and $g'(0) = g'(2\pi) = 2\pi^2$, so g would be in such space $\mathcal{X}$. However, $Lg(x) = 6x - 6\pi$, which does not satisfy the boundary conditions (in particular, $Lg(0) = -6\pi \neq 6\pi = Lg(2\pi)$. Thus any attempt of producing a fixed point Φ_g via iterations starting from $u_0 = 0$ would fail, since the iterates would escape the space $C(\mathbb{T}, \mathcal{X})$ after a finite number of steps.

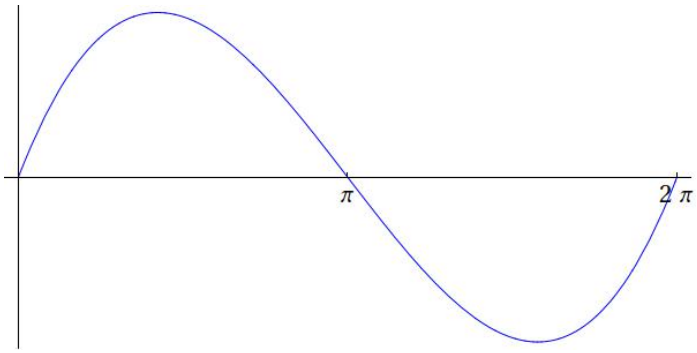


Figure 2.7: Plot of a periodic initial condition $g(x) = x(x-\pi)(x-2\pi)$..

For the rest of this section we try to obtain existence and uniqueness results for fixed points of $\hat{\Phi}_{\hat{g}}$, defined in (2.68). We recall that $C(\mathbb{T}, h^\infty(\mathbb{Z}))$ is a Fréchet space with pseudonorms

$$\|\hat{u}\|_{M,T} = \sup_{0\leq t\leq T} \|\hat{u}\|_M = \sup_{0\leq t\leq T} \sum_{k\in\mathbb{Z}} k^{2M} |\hat{u}(t)_k|^2.$$

Theorem 13 (Fixed points of $\hat{\Phi}_{\hat{g}}$: existence and convergence in $h^\infty(\mathbb{Z})$). *Let p be a polynomial of degree m and $\hat{g} \in h^\infty(\mathbb{Z})$. Suppose that there is a constant D such that, for all $M \in \mathbb{N}$,*

$$\|\hat{g}\|_M \leq D^M.$$

Then the sequence $\hat{u}^{(n)} = \hat{\Phi}_{\hat{g}}^n(0)$ converges in $C(\mathbb{T}, h^\infty(\mathbb{Z}))$ to a fixed point $\hat{u}^*$ of $\hat{\Phi}_{\hat{g}}$.

Proof. Without loss of generality assume that $p(0) = 0$, so that there exists a constant C, depending on p only, such that

$$|p(ik)| \leq C|k|^m, \quad \text{for all } k \in \mathbb{Z}. \tag{2.69}$$

We observe that $\hat{u}^{(0)}(t)_k = 0$, $\hat{u}^{(1)}(t)_k = \hat{g}_k$, $\hat{u}^{(2)}(t)_k = \hat{\Phi}_{\hat{g}}(\hat{g})(t)_k = \hat{g}_k + p(ik)t\hat{g}_k$, and, in general, we have

$$\hat{u}^{(n)}(t)_k = \sum_{\ell=0}^{n-1} \frac{p(ik)^\ell t^\ell}{\ell!} \hat{g}_k, \tag{2.70}$$

which can be proven by induction on n. Next consider $\hat{v}^{(\ell)} : \mathbb{T} \times \mathbb{Z} \to \mathbb{C}$ given by $\hat{v}^{(\ell)}(t)_k = \frac{p(ik)^\ell t^\ell}{\ell!} \hat{g}_k$. Since $\hat{g} \in h^\infty(\mathbb{Z})$, we get the following bounds for any $t \in \mathbb{T}$ and $M \in \mathbb{N}$,

$$\sum_{k\in\mathbb{Z}} k^{2M} |\hat{v}^{(\ell)}(t)_k|^2 = \sum_{k\in\mathbb{Z}} k^{2M} \left| \frac{p(ik)^\ell t^\ell}{\ell!} \hat{g}_k \right|^2 \tag{2.71a}$$

$$\leq \sum_{k\in\mathbb{Z}} k^{2M} \frac{(C|k|^m)^{2\ell} t^{2\ell}}{(\ell!)^2} |\hat{g}_k|^2 \tag{2.71b}$$

$$= \frac{(Ct)^{2\ell}}{(\ell!)^2} \sum_{k\in\mathbb{Z}} k^{2M+2m\ell} |\hat{g}_k|^2 = \frac{(Ct)^{2\ell}}{(\ell!)^2} \|\hat{g}\|^2_{M+m\ell}, \tag{2.71c}$$

where (2.71b) is justified by (2.69).

Therefore, it follows that $\hat{v}^{(\ell)} \in C(\mathbb{T}, h^\infty(\mathbb{Z}))$ and moreover, for any T, M we have

$$\|\hat{v}^{(\ell)}\|_{M,T} \leq \frac{(CT)^\ell}{\ell!} \|\hat{g}\|_{M+m\ell}. \tag{2.72}$$

We can also see that $\hat{u}^{(n)} = \sum_{\ell=0}^{n-1} \hat{v}^{(\ell)}$. Our next step is to prove that the series $\sum \hat{v}^{(\ell)}$ is absolutely convergent in $C(\mathbb{T}, h^\infty(\mathbb{Z}))$. For any $M, T \in \mathbb{N}$ we have that

$$\sum_{n=0}^{\infty} \|\hat{v}^{(\ell)}\|_{M,T} \leq \sum_{\ell=0}^{\infty} \frac{(CT)^\ell}{\ell!} \|\hat{g}\|_{M+m\ell} \tag{2.73a}$$

$$\leq \sum_{\ell=0}^{\infty} \frac{(CT)^\ell}{\ell!} D^{M+m\ell} \tag{2.73b}$$

$$= CD^M \sum_{\ell_0}^{\infty} \frac{(TD^m)^\ell}{\ell!} = CD^M e^{TD^m} < \infty, \tag{2.73c}$$

where (2.73a) is justified by (2.72) and (2.73b) is justified by the growth bounds on the pseudonorms of $\hat{g}$.

Since $\sum \hat{v}^{(\ell)}$ is absolutely convergent for any pseudonorm $\|\cdot\|_{M,T}$, it follows that $\hat{u}^{(n)}$ is a convergent sequence in $C(\mathbb{T}, h^\infty(\mathbb{Z}))$. Denoting by $\hat{u}^*$ its limit, we conclude by continuity of $\hat{\Phi}_{\hat{g}}$ that $\hat{u}^*$ is a fixed point of $\hat{\Phi}_{\hat{g}}$. □

Theorem 14 (**Fixed points of $\hat{\Phi}_{\hat{g}}$: existence, uniqueness and convergence in $\mathbb{C}^\mathbb{Z}$**). *Let p be a polynomial of degree m and $\hat{g} \in \mathbb{C}^\mathbb{Z}$.*

1. *For any $\hat{u}^{(0)} \in C(\mathbb{T}, \mathbb{C}^\mathbb{Z})$, the sequence $\hat{\Phi}^n_{\hat{g}}(\hat{u}^{(0)})$ converges in $C(\mathbb{T}, \mathbb{C}^\mathbb{Z})$ to a fixed point $\hat{u}^*$ of $\hat{\Phi}_{\hat{g}}$.*

2. $\hat{u}^*$ is the unique fixed point of $\hat{\Phi}_{\hat{g}}$ in $C(\mathbb{T}, \mathbb{C}^{\mathbb{Z}})$, and it is given by

$$\hat{u}^*(t)_k = e^{p(ik)t}\hat{g}_k. \tag{2.74}$$

Proof. The topology of $\mathbb{C}^{\mathbb{Z}}$ is that of pointwise convergence; we say that $\hat{g}^{(n)}$ converges to $\hat{g}$ if $\hat{g}^{(n)}_k \to \hat{g}_k$ for all $k \in \mathbb{Z}$. Since $\mathbb{Z}$ is countable, we can think of $\mathbb{C}^{\mathbb{Z}}$ as a Fréchet space with pseudonorms $\{\|\cdot\|_k\}_{k\in\mathbb{Z}}$ given by $\|\hat{g}\|_k = |\hat{g}_k|$. Thus $C(\mathbb{T}, \mathbb{C}^{\mathbb{Z}})$ is a Fréchet space with pseudonorms

$$\|\hat{u}\|_{k,T} = \sup_{0\le t\le T} |\hat{u}(t)_k|.$$

We first prove contraction inequalities for $\hat{\Phi}_{\hat{g}}$. Let $n \in \mathbb{N}$ and consider a pseudonorm $\|\cdot\|_{k,T}$. We have

$$\begin{aligned}
\left\|\hat{\Phi}^n_{\hat{g}}\hat{u} - \hat{\Phi}^n_{\hat{g}}\hat{v}\right\|_{k,T} &= \left\|\int_0^t \cdots \int_0^{s_{n-1}} p(ik)^n(\hat{u}-\hat{v})ds_n \dots ds_1\right\|_{k,T} && \text{(2.75a)}\\
&= \sup_{0\le t\le T}\left|\int_0^t \cdots \int_0^{s_{n-1}} p(ik)^n(\hat{u}-\hat{v})ds_n \dots ds_1\right| && \text{(2.75b)}\\
&\le \sup_{0\le t\le T}\int_0^t \cdots \int_0^{s_{n-1}} |p(ik)|^n|\hat{u}(s_n)_k - \hat{v}(s_n)_k|ds_n \dots ds_1 && \text{(2.75c)}\\
&\le \frac{|p(ik)|^nT^n}{n!}\sup_{0\le t\le T}|\hat{u}(t)_k - \hat{v}(t)_k| && \text{(2.75d)}\\
&\le \frac{|p(ik)|^nT^n}{n!}\|\hat{u}-\hat{v}\|_{k,T}. && \text{(2.75e)}
\end{aligned}$$

Next we take an arbitrary $\hat{u}^{(0)} \in C(\mathbb{T}, \mathbb{C}^{\mathbb{Z}})$ and prove that the sequence $\hat{u}^{(n)} = \hat{\Phi}^n_{\hat{g}}\hat{u}^{(0)}$ has a limit. We fix $k, T \in \mathbb{N}$ and observe that

$$\begin{aligned}
\sum_{n=0}^{\infty}\|\hat{u}^{(n+1)} - \hat{u}^{(n)}\|_{k,T} &= \sum_{n=0}^{\infty}\|\hat{\Phi}^n_{\hat{g}}\hat{u}^{(1)} - \hat{\Phi}^n_{\hat{g}}\hat{u}^{(0)}\|_{k,T}\\
&\le \sum_{n=0}^{\infty}\frac{|p(ik)|^nT^n}{n!}\|\hat{u}^{(1)} - \hat{u}^{(0)}\|_{k,T}\\
&= e^{|p(ik)|T}\|\hat{u}^{(1)} - \hat{u}^{(0)}\|_{k,T} < \infty.
\end{aligned}$$

Repeating the reasoning of previous proofs (cf. Theorem 7), we then conclude that $(\hat{u}^{(n)})$ is a Cauchy sequence and thus it must converge to some limit $\hat{u}^*$, which must be (by continuity) a fixed point of $\hat{\Phi}_{\hat{g}}$. This proves the first claim.

To prove uniqueness, let $\hat{u}^*$, $\hat{v}^* \in C(\mathbb{T}, \mathbb{C}^{\mathbb{Z}})$ be fixed points of $\hat{\Phi}_{\hat{g}}$. Take any pseudonorm $\|\cdot\|_{k,T}$ and let n be large enough such that $\frac{|p(ik)|^nT^n}{n!} < 1$. Then

$$\|\hat{u}^* - \hat{v}^*\|_{k,T} = \|\hat{\Phi}^n_{\hat{g}}(\hat{u}^*) - \hat{\Phi}^n_{\hat{g}}(\hat{v}^*)\|_{k,T} \le \frac{|p(ik)|^nT^n}{n!}\|\hat{u}^* - \hat{v}^*\|_{k,T}, \tag{2.76}$$

which implies $\|\hat{u}^* - \hat{v}^*\|_{k,T} = 0$. As k and T were arbitrary, we conclude that $\hat{u}^* = \hat{v}^*$.

Finally, to prove (2.74), we let $\hat{u}^*(t)_k = e^{p(ik)t}\hat{g}_k$ and compute

$$\hat{\Phi}_{\hat{g}}\hat{u}^*(t)_k = \hat{g}_k + \int_0^t p(ik)e^{p(ik)s}\hat{g}_k ds = \hat{g}_k + \Big[e^{p(ik)s}\hat{g}_k\Big]_{s=0}^{s=t}$$
$$= \hat{g}_k + e^{p(ik)t}\hat{g}_k - \hat{g}_k = e^{p(ik)t}\hat{g}_k.$$

Therefore, $\hat{u}^*$ is a fixed point of $\hat{\Phi}_{\hat{g}}$. By uniqueness, it follows that any sequence of iterates $\hat{\Phi}^n_{\hat{g}}\hat{u}^{(0)}$ must converge to $\hat{u}^*$, which concludes the proof. □

Proposition 2.7.6. *Let p be a polynomial of degree m and $\hat{g} \in h^\infty(\mathbb{Z})$. Suppose that there is a constant C such that, for all $k \in \mathbb{Z}$, $\mathrm{Re}(p(ik)) \leq C$. Then $\hat{u}^*$ defined as in (2.74) is in $C(\mathbb{T}, h^\infty(\mathbb{Z}))$.*

Proof. For any pseudonorm $\|\cdot\|_{T,M}$ we have

$$\|\hat{u}^*\|^2_{M,T} = \sup_{0\leq t\leq T}\sum_{k\in\mathbb{Z}} k^{2M}|\hat{u}^*(t)_k|^2 \tag{2.77a}$$
$$= \sup_{0\leq t\leq T}\sum_{k\in\mathbb{Z}} k^{2M}|e^{p(ik)t}|^2|\hat{g}_k|^2 \tag{2.77b}$$
$$\leq \sup_{0\leq t\leq T}\sum_{k\in\mathbb{Z}} k^{2M}e^{2Ct}|\hat{g}_k|^2 \tag{2.77c}$$
$$= e^{2CT}\sum_{k\in\mathbb{Z}} k^{2M}|\hat{g}_k|^2 = e^{2CT}\|\hat{g}\|_M < \infty, \tag{2.77d}$$

where (2.77c) is justified by the bound on the real part of $p(ik)$. Thus $\hat{u}^* \in C(\mathbb{T}, h^\infty(\mathbb{Z}))$, as we wanted to prove. □

2.8 Discussion

The goal of this chapter was to study linear evolution problems and present some useful notions and concepts. As we have seen, the notion of Fréchet space appears to be fundamental in our framework. In other words, the topology of the underlying space is induced by a *family of pseudonorms* instead of just a norm. This may happen for two reasons related to some type of *unboundedness*; first, the space $\mathcal{X}$ may correspond to functions in an unbounded domain, such as $\mathcal{X} = C(\mathbb{R})$; second, the time domain may be itself unbouded, $\mathbb{T} = [0, \infty)$. We also remark that the framework of Fréchet spaces was implicitly present in the work by Tucker and Zucker; however, it had not been discussed in detail before.

As we mentioned, linear differential operators are continuous in the Fréchet space of infinitely differentiable functions. By considering analytic initial conditions, we saw how the Cauchy-Kowalevski theory allows for finding of analytic solutions. In the case of the transport equation, we can obtain contraction inequalities, which can be used to prove convergence of fixed points. As a next step, one may look as a more general operator $L : \mathcal{X} \to \mathcal{X}$ using higher-order derivatives, for example with bounds of the form

$$\|Lu\|_{T,X,k} \leq C\|\partial_x^\ell u\|_{T,X,k} \leq C\|u\|_{T,X,k+\ell}. \tag{2.78}$$

We observe that analyticity of the initial condition g is not enough to ensure existence of solutions. A counterexample is given by the heat equation $u_t = u_{xx}$ with initial condition $g(x) = \frac{1}{1-x}$. Even though g is analytic near zero, the solution fails to be analytic at a neighborhood of the origin

(see [ES98]). Thus, the general case may require different tools such as explicit bounds on the pseudonorms of g.

We also tested an approach to obtain fixed points using the Fourier Transform. In this case, analyticity is no longer a requirement; however, we had to restrict our attention to functions with a special decay at infinity, such as the Schwarz space $\mathcal{S}(\mathbb{R})$. We have seen that the usual theory for solving linear partial differential equations in this setting can be reformulated as existence and uniqueness of fixed points in $C(\mathbb{T}, \mathcal{S}(\mathbb{R}))$. Moreover, by allowing a coarser topology in the frequency space (corresponding to the space $C(\mathbb{T}, C^\infty(\mathbb{R}))$), we also obtained convergence of iterations to the fixed point. However, there remained a gap in trying to prove convergence in the finer topology, which is left as an open problem.

It is possible to use continuity to extend the results from the Schwarz space to finer spaces, such as $L^2(\mathbb{R})$ or $H^s(\mathbb{R})$, using boundedness of the solution operators; see for example [Rau91]. We chose not to study such results in the fixed point framework for two reasons. First, we are mostly interested in spaces of continuous functions, since we want to talk about computability properties. Second, the finer spaces still require some special behaviour at infinity (at least integrability), which leaves out many interesting functions (such as polynomials).

As a next step, one could wonder if the Fourier Transform approach could be extended to C^∞-functions. We know that the solution of the initial value problem $\frac{du}{dt} = p(\partial_x)u$ with $u(0) = g \in \mathcal{S}(\mathbb{R})$ can be given by

$$u(t,x) = \mathcal{F}^{-1} e^{p(i\xi)t} \mathcal{F} g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(x-y)\xi + p(i\xi)t} g(y) dy d\xi;$$

can we make sense of this construction for $g \in C^\infty(\mathbb{R})$? For example, one could approximate the Fourier Transforms and replace the improper integrals by finite integrals such as $\int_{-n}^{n} dx$, and then study the limit $n \to \infty$. We leave this approach as an open question.

Chapter 3

Semantics of Analog Systems

In this chapter we develop a model of analog computation that will be studied extensively in our thesis. Essentially, this model is a generalization of Shannon's General Purpose Analog Computer (GPAC) [Sha41]. In the Shannon GPAC channels carry real-valued streams; in our model, channel values can lie in a general complete metric vector space $\mathcal{X}$, such as a Banach or a Fréchet space. This allows us to, among other things, establish a framework dealing with functions of more than one variable.

We begin by presenting the Shannon GPAC, its semantics and some basic examples of generable functions. We also present Shannon's characterization theorem in terms of differentially algebraic functions. We then move to a model called $\mathcal{X}$-GPAC, whose channels carry function-valued streams. We define semantics of an $\mathcal{X}$-GPAC using a notion of *quasi-well-posedness* and obtain a characterization of the $\mathcal{X}$-GPAC-generable functions. Afterwards, we consider a multityped GPAC and present various modular operations such as module derivation and channel contraction.

Let us comment on the original content of this chapter. The idea of considering function-valued streams (as opposed to real-valued streams) is arguably our biggest contribution to this field of research. Consequently, the $\mathcal{X}$-GPAC (Definition 3.4.4) is an original model and Theorem 17 extends Shannon's characterization into the realm of functions of more than one variable and partial differential equations in a novel way. The notion of quasi-well-posedness (Definition 3.4.7) is an original adaptation of the notion of well-posedness existing in literature (see [Had52, CH53]). The last part of the chapter, devoted to a multityped GPAC, provides a new and perhaps promising direction of research; unfortunately it is not pursued in much detail and only some basic results are proved. We believe that Lemma 3.9.4, which relates contractive operators with channel contractions and motivates one of the findings in [Jam12], is the most interesting result in this part.

3.1 The Shannon GPAC

We start by presenting the Shannon GPAC, originally introduced in [Sha41], and later improved by [PE74, LR87, GC03].

The construction of an analog system presupposes the notion of channels, which carry information, and modules, which operate on channels. In the original construction of Shannon, there is only one channel type:

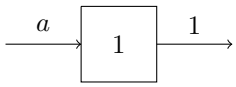
- *real-valued stream* channels, which carry a real-valued stream $a \in C^1(\mathbb{T}, \mathbb{R})$.

In contrast to the previous chapter, we shall only deal with bounded time $\mathbb{T} = [0, T]$, where $T \in \mathbb{R}^+$ denotes the final time. The case where $\mathbb{T} = [0, \infty)$ (unbounded time) can be treated in a similar manner. We recall that $C^1(\mathbb{T}, \mathbb{R})$ denotes the class of continuously differentiable functions from $\mathbb{T}$ to $\mathbb{R}$.

Each module has zero, one or more input channels, and must have a single output channel. In the original construction of Shannon, there are only four types of modules, which we now present.

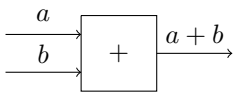
Definition 3.1.1 (**Basic Shannon modules**)**.** The *basic Shannon modules* are defined as follows:

- the *Shannon constant* module has one input and one output. For input a, it outputs the constant stream $b \equiv 1$;
- The *Shannon adder* module has two inputs and one output. For inputs a and b, it outputs the sum $a + b$;
- for each $k \in \mathbb{R}$, we define a *Shannon scalar multiplier* module with one input and one output. For input a, it outputs the product ka;
- the *Shannon integrator* module has two inputs and one output. For inputs a and b, it outputs the Lebesgue-Stieltjes integral $k + \int a db$, where $k \in \mathbb{R}$ is an initial setting.



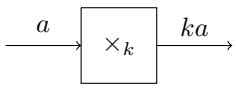
$$1 : C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$1(a)(t) = 1$$



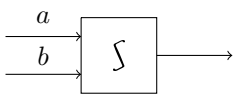
$$+ : C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$+(a,b)(t) = a(t) + b(t)$$



$$\times_k : C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$\times_k(a)(t) = k \cdot a(t)$$



$$\textstyle\int : C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$\textstyle\int(a,b)(t) = k + \int_0^t a(s)db(s)$$

Figure 3.1: The four basic Shannon modules.

Remark 3.1.2. The integrator module is well defined, since the Lebesgue-Stieltjes integral is well defined for continuous integrand and continuously differentiable integrator. In other words, for any $a, b \in C^1([0,T],\mathbb{R})$, the expression $\int_0^t a(s)db(s)$ defines a function in $C^1([0,T],\mathbb{R})$.

Remark 3.1.3. We also introduce the symbol '$\int$' to denote the operator associated with the integrator module, in order to differentiate from the actual integral; we can then write $\int(a,b) = k + \int a db$.

Remark 3.1.4. The reader familiar with computable analysis may wonder if the space of multiplication modules is too broad and if, instead, we should consider only the multiplication modules associated with *computable* real numbers $k \in \mathbb{R}$. Our choice to allow any real number is compatible with the existing literature on the subject. Later on, in Chapter 5, we will be interested in studying concrete computability, in which case we must enforce computability of the constants involved.

Definition 3.1.5 (**Shannon GPAC**)**.** A *Shannon general purpose analog computer* (GPAC) is a network built with the four Shannon basic modules (constants, adders, multipliers and integrators) and connections between their inputs and outputs, with the following restrictions:

- the only connections allowed are between an output and an input;
- each input may be connected to either zero or one output;

Remark 3.1.6. Of course, there can be cycles in a GPAC, for example, an output connected to an input of its own module. This is called *feedback* and it is fundamental to develop an interesting theory.

Example 3.1.7. A simple example of a Shannon GPAC can be seen in Figure 3.2.

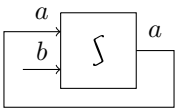


Figure 3.2: A GPAC for computing the exponential function.

This GPAC has only one module, which is an integrator module, and only one connection, between its output channel and one of its input channels.

Once we have a GPAC, such as the one in the previous example, we would like to study what function or tuple of functions, if any, is *generated* by that system (in other words, we want to attribute semantics to each GPAC). In other to do that, we define the notion of operator induced by a GPAC.

Definition 3.1.8 (**Shannon GPAC induced operator**)**.** Let $\mathcal{G}$ be a Shannon GPAC. We define:

- the *constant space* of $\mathcal{G}$ is the cartesian product of the spaces associated with each constant occuring in an integrator. This constant space can be written as $\mathcal{C} = \mathbb{R}^p$, for some $p \geq 0$;
- the *proper input space* of $\mathcal{G}$ is the cartesian product of the spaces associated with each unconnected input channel. This input space can be written as $\mathcal{I} = C^1([0,T],\mathbb{R})^q$, for some $q \geq 0$;
- the *proper output space* of $\mathcal{G}$ is the cartesian product of the spaces associated with each unconnected output channel. This output space can be written as $\mathcal{O} = C^1([0,T],\mathbb{R})^m$, for some $m \geq 0$;
- the *mixed space* of $\mathcal{G}$ is the cartesian product of the spaces associated with each channel which connects an input with an output. This mixed space can be written as $\mathcal{M} = C^1([0,T],\mathbb{R})^r$, for some $r \geq 0$;

- the *induced operator* of $\mathcal{G}$ is the function

$$\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}, \quad \Phi(\mathbf{k}, \mathbf{a}^I, \mathbf{a}^M) = (\tilde{\mathbf{a}}^M, \mathbf{a}^O), \tag{3.1}$$

  where $\mathbf{k}$ is a vector of scalars and each $\mathbf{a}$ is a vector of stream channels. Moreover, each of the components in the codomain of Φ is given by the module with which it is associated. In other words, for each a_i component of either $\tilde{\mathbf{a}}^M$ or $\mathbf{a}^O$,

  - if a_i is associated with the output channel of a constant, then $a_i = 1(b)$, where b is the component in $\mathbf{a}^I$ or $\mathbf{a}^M$ associated with the input channel of that module;
  - if a_i is associated with the output channel of an adder, then $a_i = +(b_1, b_2)$, where b_1 and b_2 are the components in $\mathbf{a}^I$ or $\mathbf{a}^M$ associated with the input channels of that module;
  - if a_i is associated with the output channel of a scalar multiplier, then $a_i = \times_k(b)$, where b is the component in $\mathbf{a}^I$ or $\mathbf{a}^M$ associated with the input channel of that module, and k is its corresponding multiplication factor;
  - if a_i is associated with the output channel of an integrator, then $a_i = \int(b_1, b_2)$, where b_1, b_2 are the components in $\mathbf{a}^I$ or $\mathbf{a}^M$ associated with the input channels of that module, and k is the component in $\mathbf{k}$ associated with the initial setting of that module.

Example 3.1.9. The above definition may seem verbose or even pedantic. For the sake of exposition, let us come back to Example 3.1.7 and see what the induced operator is. In this situation, there is only one constant k, one unconnected input channel b, and one mixed channel a. There are no unconnected output channels. Thus, the vector $\mathbf{a}^O$ is empty, and the other vectors only have one component, $\mathbf{k} = k$, $\mathbf{a}^I = b$ and $\mathbf{a}^M = a$. The variable $\tilde{a}$ associated with the only mixed channel must be given by the formula for the integrator (Figure 3.1). Thus, the induced operator is given by

$$\Phi : \mathbb{R} \times C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R});$$

$$\Phi(k,a,b)(t) = k + \int_0^t a(s)db(s). \tag{3.2}$$

Remark 3.1.10. A general Shannon GPAC, with proper input space $\mathcal{I}$, mixed space $\mathcal{M}$ and proper output space $\mathcal{O}$ may be represented in a diagram as in Figure 3.3. Notice that the constants (i.e. initial settings of integrators) are implicit in the GPAC diagram but are explicit in the description of the induced operator.

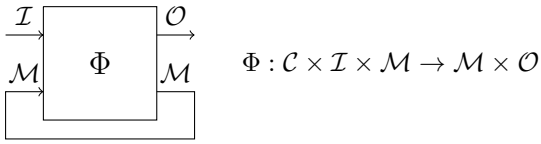


Figure 3.3: General diagram for a Shannon GPAC.

It should be clear that, once the induced operator is defined, the next task is to search for fixed points (we still need to clarify what a fixed point is, since the domain and codomain of Φ do not match in general). Not all possible constructions of a GPAC are desirable; in fact, we should only be interested in those systems for which the fixed point problem is well-posed, in the sense of the following definition.

Definition 3.1.11 (**Well-posedness of Shannon GPAC**)**.** Let $\mathcal{G}$ be a Shannon GPAC and $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}$ be its induced operator. Let U be an open subset of $\mathcal{C} \times \mathcal{I}$. We say that $\mathcal{G}$ is *well-posed* on U if

- (*existence*) for every $(\mathbf{k}, \mathbf{a}^I) \in U$, there exists $(\mathbf{a}^M, \mathbf{a}^O) \in \mathcal{M} \times \mathcal{O}$ such that

$$\Phi(\mathbf{k}, \mathbf{a}^I, \mathbf{a}^M) = (\mathbf{a}^M, \mathbf{a}^O); \tag{3.3}$$

- (*uniqueness*) for every $(\mathbf{k}, \mathbf{a}^I) \in U$, the tuple $(\mathbf{a}^M, \mathbf{a}^O)$ such that (3.3) holds is unique;
- (*continuity*) the map $(\mathbf{k}, \mathbf{a}^I) \mapsto (\mathbf{a}^M, \mathbf{a}^O)$, with domain U and codomain $\mathcal{M} \times \mathcal{O}$, given as the unique solution of (3.3), is continuous.

Let $\mathbf{k} \in \mathcal{C}$ and $\mathbf{a}^I \in \mathcal{I}$. We say that $\mathcal{G}$ is *well-posed* at $(\mathbf{k}, \mathbf{a}^I)$ if $\mathcal{G}$ is well-posed in some neighbourhood of $(\mathbf{k}, \mathbf{a}^I)$.

Definition 3.1.12 (**Semantics of Shannon GPAC**)**.** Let $\mathcal{G}$ be a Shannon GPAC having induced operator $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}$. Let U be an open subset of $\mathcal{C} \times \mathcal{I}$ such that $\mathcal{G}$ is well-posed on U. The *specification* of $\mathcal{G}$ on U is the (partial) function

$$\begin{array}{l} F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}; \\ F(\mathbf{k}, \mathbf{a}^I) = (\mathbf{a}^M, \mathbf{a}^O), \end{array} \tag{3.4}$$

whose domain is U and where $(\mathbf{a}^M, \mathbf{a}^O)$ is given by (3.3). We also say that $\mathcal{G}$ *generates* F on $U = \mathrm{dom}(F)$.

A function $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$ is Shannon GPAC-*generable* if its domain is an open set, and there exists a Shannon GPAC $\mathcal{G}$ such that $\mathcal{G}$ is well-posed on the domain of F and F is the specificaton of $\mathcal{G}$.

Example 3.1.13. Returning to Example 3.1.7, whose induced operator is given by (3.2); the fixed point equation (3.3) becomes

$$k + \int_0^t a(s) db(s) = a(t); \tag{3.5}$$

since $a, b \in C^1([0, T], \mathbb{R})$ we can differentiate both sides to obtain

$$a'(t) = a(t) b'(t); \tag{3.6}$$

this is a linear ODE whose solution can be easily seen to be (note that (3.5) implies $a(0) = k$)

$$a(t) = k e^{b(t) - b(0)}; \tag{3.7}$$

we then conclude that this GPAC is well-posed on the whole space $\mathcal{C} \times \mathcal{I} = \mathbb{R} \times C^1([0, T], \mathbb{R})$, for any $T \in \mathbb{R}^+$, and generates the function

$$F : \mathbb{R} \times C^1([0, T], \mathbb{R}) \to C^1([0, T], \mathbb{R}); \quad F(k, b)(t) = k e^{b(t) - b(0)}. \tag{3.8}$$

In the case that $k = 1$ and b is linear time, that is, $b(t) = t$, the output of F is the exponential function $F(1, b) : t \mapsto e^t$.

Remark 3.1.14. The function $t \mapsto t$ will appear frequently in this thesis; we shall refer to it as *linear time* and denote this function by $\mathbf{t}$.

We will be interested in generalizing the Shannon GPAC to functions of more than one variable in this thesis, but before that, we should present some results that illustrate the power of GPACs.

Lemma 3.1.15 (**Basic operations on GPACs**).

- *Let* $F_1 : \mathcal{C}_1 \times \mathcal{I}_1 \rightharpoonup \mathcal{M}_1 \times \mathcal{O}_1$ *and* $F_2 : \mathcal{C}_2 \times \mathcal{I}_2 \rightharpoonup \mathcal{M}_2 \times \mathcal{O}_2$ *be Shannon GPAC-generable; write*

$$F_1(\mathbf{cst}_1, \mathbf{in}_1) = (\mathbf{mix}_1, \mathbf{out}_1); \quad F_2(\mathbf{cst}_2, \mathbf{in}_2) = (\mathbf{mix}_2, \mathbf{out}_2),$$

  whenever F_1 *and* F_2 *are defined; then the* parallel composition $(F_1, F_2) : \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{I}_1 \times \mathcal{I}_2 \rightharpoonup \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{O}_1 \times \mathcal{O}_2$*, given by*

$$(F_1, F_2)(\mathbf{cst}_1, \mathbf{cst}_2, \mathbf{in}_1, \mathbf{in_2}) = (\mathbf{mix}_1, \mathbf{mix}_2, \mathbf{out}_1, \mathbf{out}_2),$$

  whenever F_1 *and* F_2 *are both defined, is Shannon GPAC-generable.*

- *Let* $F_1 : \mathcal{C}_1 \times \mathcal{I}_1 \rightharpoonup \mathcal{M}_1 \times \mathcal{O}_1$ *and* $F_2 : \mathcal{C}_2 \times \mathcal{I}_2 \rightharpoonup \mathcal{M}_2 \times \mathcal{O}_2$ *be Shannon GPAC-generable; assume that*

$$\mathcal{I}_2 = \mathcal{M}_1 \times \mathcal{O}_1$$

  (that is, the mixed and output channels of F_1 *agree with the input channels of* $\mathcal{I}_2$*); write*

$$F_1(\mathbf{cst}_1, \mathbf{in}_1) = (\mathbf{mix}_1, \mathbf{out}_1);$$

$$F_2(\mathbf{cst}_2, \mathbf{mix}_1, \mathbf{out}_1) = (\mathbf{mix}_2, \mathbf{out}_2),$$

  whenever F_1 *and* F_2 *are defined; then the* serial composition $F_2 \circ F_1 : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$*, given by*

$$\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2; \quad \mathcal{I} = \mathcal{I}_1; \quad \mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{O}_1; \quad \mathcal{O} = \mathcal{O}_2;$$

$$F_2 \circ F_1(\mathbf{cst}_1, \mathbf{cst}_2, \mathbf{in}_1) = (\mathbf{mix}_1, \mathbf{mix}_2, \mathbf{out}_1, \mathbf{out}_2),$$

  is Shannon GPAC-generable.

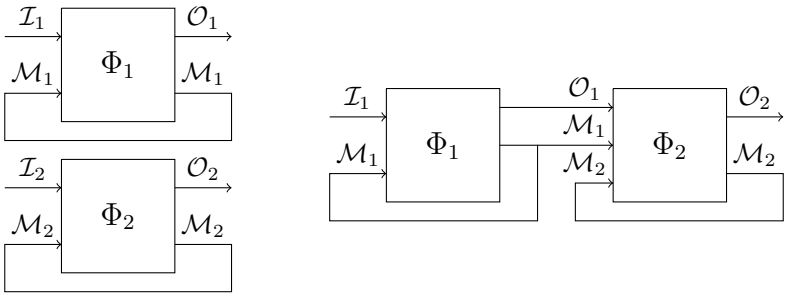


Figure 3.4: Parallel and serial composition of Shannon GPACs.

Proof. Just observe that, given Shannon GPACs which generate F_1 and F_2, we can combine them to obtain a Shannon GPAC which generates the parallel composition (F_1, F_2) or the serial composition $F_2 \circ F_1$, as in Figure 3.4. We leave the details to the reader. □

Remark 3.1.16. Mixtures of parallel and serial composition are possible; for example, one may imagine that some input channels connect only to $\mathcal{G}_1$, or only to $\mathcal{G}_2$, or to both; and that some

mixed and output channels of $\mathcal{G}_1$ connect as input channels of $\mathcal{G}_2$, while others do not. Each of these possible mixtures can in turn be abstracted in a diagram similar to those in Figure 3.4.

Lemma 3.1.17 (**Normal form Lemma**).

(a) Let $\mathcal{G}$ be a Shannon GPAC generating a function $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$. Assume $\mathcal{I} = C^1([0,T],\mathbb{R})^p$ and $\mathcal{M} \times \mathcal{O} = C^1([0,T],\mathbb{R})^q$. Then all tuples of the form $(\mathbf{in}, F(\mathbf{cst},\mathbf{in})) = (\mathbf{in},\mathbf{mix},\mathbf{out})$ satisfy a system of q differential equations in $q+p$ variables

$$\sum_{i=0}^{q+p}\sum_{j=1}^{q+p} a_{ijk} z_i z_j' = 0; \quad k = 1,\ldots,q, \tag{3.9}$$

where a_{ijk} are real constants, $z_0 \equiv 1$, the variables $z_1,\ldots,z_p$ correspond to the input channels and the variables $z_{p+1},\ldots,z_{p+q}$ correspond to the mixed and output channels. This system can also be rearranged in the form

$$A(\mathbf{x},\mathbf{y})\cdot\mathbf{y}' = B(\mathbf{x},\mathbf{y})\cdot\mathbf{x}', \tag{3.10}$$

where $\mathbf{x} = (z_1,\ldots,z_p)$, $\mathbf{y} = (z_{p+1},\ldots,z_{p+q})$, A is a $q\times q$ matrix, B is a $q\times p$ matrix, and the coefficients of A and B are linear in $1,\mathbf{x},\mathbf{y}$.

(b) Conversely, if systems (3.9) or (3.10) have a well-posed solution for initial conditions

$$z_{p+1}(0) = C_1,\ldots,z_{p+q}(0) = C_q,$$

and inputs $z_1,\ldots,z_p$ in some open set $U \subseteq \mathbb{R}^q \times C^1([0,T],\mathbb{R})^p$, then the map

$$(C_1,\ldots,C_q,z_1,\ldots,z_p) \mapsto (z_{p+1},\ldots,z_{p+q})$$

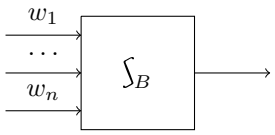
is the projection of a GPAC-generable function

$$F : \mathcal{C}\times\mathcal{I} \rightharpoonup \mathcal{M}\times\mathcal{O}$$

onto some of its mixed and output channels.

The main idea of proving the lemma above is by reducing all modules in the GPAC construction to a single module type.

Definition 3.1.18 (**Integral-matrix module**). For each $(n+1)\times n$ matrix B, we define an *integral-matrix module* with n inputs and one output. For inputs $w_1,\ldots,w_n$, it outputs the stream $k + \sum_{i=0}^{n}\sum_{j=1}^{n} b_{ij}\int w_i dw_j$, where k is an initial setting and $w_0 \equiv 1$.



$$\int_B : C^1([0,T],\mathbb{R})^n \to C^1([0,T],\mathbb{R})$$
$$\int_B (w_1,\ldots,w_n)(t) = k + \sum_{i=0}^{n}\sum_{j=1}^{n} b_{ij}\int_0^t w_i(s)w_j'(s)ds$$

Figure 3.5: The integral-matrix module.

Proof. For part (a), we observe that each of the four Shannon basic modules are special cases of the integral-matrix module. For example, the identity

$$a(t) + b(t) = a(0) + b(0) + \int_0^t a'(s)ds + \int_0^t b'(s)ds$$

implies that the adder module can be expressed as the integral-matrix module with 2 inputs, initial setting $a(0) + b(0)$ and matrix B such that $b_{01} = b_{02} = 1$ and $b_{ij} = 0$ for the other coefficients, with the correspondence $w_0 \equiv 1$, $w_1 \equiv a$, $w_2 \equiv b$. Similarly, the expression defining the integrator module,

$$k + \int_0^t a(s)b'(s)ds$$

implies that the integrator module can be expressed as the integral-matrix module with two inputs, initial setting k and matrix B such that $b_{12} = 1$ and $b_{ij} = 0$ for the other coefficients, with the correspondence $w_0 \equiv 1$, $w_1 \equiv a$, $w_2 \equiv b$. We leave the constant and scalar multiplier modules to the reader and summarize the results in the following table.

Type of module	Expression	B matrix	initial setting
Constant	1	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	1
Adder	$a + b$	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$	$a(0) + b(0)$
Scalar multiplier	ka	$\begin{bmatrix} k \\ 0 \end{bmatrix}$	$ka(0)$
Integrator	$k + \int adb$	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$	k

Let $\mathcal{G}$ be a Shannon GPAC in the conditions of the lemma. Let us label the proper input channels with $z_1, \ldots, z_p$ and the remaining (proper output and mixed) channels with $z_{p+1}, \ldots, z_{p+q}$. Let us also introduce $z_0 \equiv 1$ for ease of notation. By writing each of the modules of $\mathcal{G}$ in the integral-matrix formulation, we obtain

$$z_k(t) = C_k + \sum_{i=0}^{p+q} \sum_{j=1}^{p+q} b_{ijk} \int_0^t z_i(s) z_j'(s) ds, \quad k = p+1, \ldots, p+q \tag{3.11}$$

where C_k may depend on the initial values of some z_i or on an initial setting. Differentiating in time, we get

$$z_k'(t) = \sum_{i=0}^{p+q} \sum_{j=1}^{p+q} b_{ijk} z_i(t) z_j'(t); \tag{3.12}$$

rearranging this expression gives us (3.9). Of course, we can switch between (3.9) and (3.10), using

the conversions

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \quad \begin{bmatrix} -\mathbf{B}(\mathbf{x},\mathbf{y}) & \mathbf{A}(\mathbf{x},\mathbf{y}) \end{bmatrix}_{k=1,j=1}^{q,p+q} = \left[\sum_{i=0}^{p+q} a_{ijk} z_i \right].$$

For the converse direction (b), we start from (3.9) and write it in the form (3.12); after integrating both sides we get the integral-matrix formulation (3.11). We can obtain the integral-matrix formulation as a composition of the Shannon basic modules, and thus we can construct a GPAC that implements (3.11), possibly with extra initial settings and extra mixed (auxiliary) channels (see Figure 3.6 for an example with $n = 2$ channels).

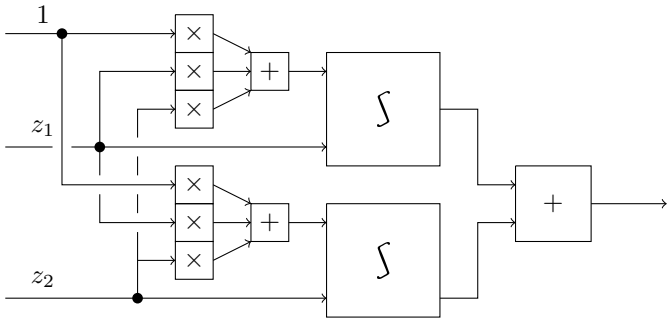


Figure 3.6: Reduction of the integral-matrix module via the Shannon basic modules, for $n = 2$; the channel labelled 1 can be obtained as the output of the constant module; the coefficients b_{ij} of the B matrix correspond to the scalar multiplier modules, but are omitted for simplicity.

Since (3.9) is, by assumption, well-posed on some open set $U \subseteq \mathbb{R}^q \times C^1([0,T],\mathbb{R})^p$ of initial conditions and inputs and the extra channels are uniquely continuously determined by the z_i, we get that the GPAC is also well-posed on U and generates a function $F : \mathbb{R}^q \times C^1([0,T],\mathbb{R})^p \rightharpoonup \mathcal{M} \times \mathcal{O}$. The channels $z_{p+1}, \ldots, z_{p+q}$ will correspond to some of the channels in $\mathcal{M} \times \mathcal{O}$ and can be obtained with a suitable projection. □

Remark 3.1.19. When $p = 1$ and $\mathbf{x} = \mathbf{t}$, that is, there is only one input channel x, which is given by $x(t) = t$ (cf. Remark 3.1.14), equation (3.10) becomes

$$A(t,\mathbf{y}) \cdot \mathbf{y}' = B(t,\mathbf{y}), \tag{3.13}$$

where $\mathbf{y} = (y_1, \ldots, y_q)$ can be regarded as a tuple of functions of t, A is a $q \times q$ matrix, B is a $q \times 1$ vector, and the coefficients of A and B are linear in $1, t, y_1, \ldots, y_q$. This is the form that appears in [PE74] and is widely used as the definition of GPAC-generability, instead of Definition 3.1.12.

Remark 3.1.20. We must remark that (3.9) is nonlinear in the variables $z_1, \ldots, z_{p+q}$; in fact, a way of understanding (3.9) is realizing that GPACs have the ability to generate pairwise products of variables. For example, the relation $z_1 = z_2 z_3$ can be written as $z_1' - z_2 z_3' - z_3 z_2' = 0$, which is in the form of (3.9). Another example is the relation $z_1 = z_2^2$, which can be written as $z_1' - 2 z_2 z_2' = 0$, again in the form of (3.9). Using intermediate variables and system of equations we can easily devise ways to obtain relations between products of higher power, like $z_1 = z_2 z_3 z_4$ or $z_1 = z_2^4$. In fact, any expression which is *polynomial* in its variables should be 'generable' (in some sense) by a GPAC. This is the fundamental result in the study of the Shannon GPAC, and is formulated in Theorems 15 and 16 below. Their proofs use parts (a) and (b), respectively, of the Normal Form Lemma (Lemma 3.1.17).

Definition 3.1.21 (**Differentially algebraic function**). A function $f : [0,T] \to \mathbb{R}$ is said to be *differentially algebraic* if there exists $k \in \mathbb{N}$ and a polynomial P in $k+2$ variables such that $f \in C^k([0,T])$ and

$$P(t, f(t), f'(t), \ldots, f^{(k)}(t)) = 0, \quad \text{for all } t \in [0,T]. \tag{3.14}$$

Theorem 15 (**GPAC-generability implies differential algebraicity**). *Let $\mathcal{G}$ be a Shannon GPAC with one input channel (that is, $\mathcal{I} = C^1([0,T],\mathbb{R})$), well-posed in some open set $U \subseteq \mathcal{C} \times \mathcal{I}$, and let $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$ be its specification. Then, for any $\mathbf{k} \in \mathcal{C}$ such that $(\mathbf{k}, \mathbf{t}) \in U$ (that is, $\mathcal{G}$ is well-posed at $(\mathbf{k}, \mathbf{t})$), we have that $F(\mathbf{k}, \mathbf{t})$ is a tuple of differentially algebraic functions (that is, each component $F_i(\mathbf{k}, \mathbf{t})$ of $F(\mathbf{k}, \mathbf{t})$ is differentially algebraic).*

Example 3.1.22. Returning to example 3.1.7, which was seen to specify the functional $F(k,b)(t) = ke^{b(t)-b(0)}$, we can then conclude using Theorem 15 that $f(t) = e^t = F(1, \mathbf{t})(t)$ is differentially algebraic. Of course, one might simply directly verify that $f(t) = e^t$ is differentially algebraic since it satisfies $f'(t) - f(t) = 0$.

Theorem 16 (**Differential algebraicity implies GPAC-generability**). *Let $f : [0,T] \to \mathbb{R}$ be differentially algebraic. Let $k \in \mathbb{N}$ and P be a polynomial in $k+2$ variables such that $f \in C^k([0,T],\mathbb{R})$ and* (3.14) *holds. Assume in addition that $k \geq 1$ and that*

$$P(t, g(t), g'(t), \ldots, g^{(k)}(t)) = 0, \quad g(0) = y_0, g'(0) = y_1, \ldots, g^{(k-1)}(0) = y_{k-1},$$

is a well-posed problem in $C^k([0,T],\mathbb{R})$ and on a open subset U of $\mathbb{R}^k$ containing $(f(0), \ldots f^{(k-1)}(0))$. Then $f(t) = F_i(f(0), f'(0), \ldots, f^{(k-1)}(0), \mathrm{id})(t)$, where F_i is a component of a Shannon GPAC-generable function $F : \mathbb{R}^k \times C^1([0,T],\mathbb{R}) \rightharpoonup \mathcal{M} \times \mathcal{O}$.

Remark 3.1.23. An original proof of Theorems 15 and 16 can be found in [Sha41], but this proof had flaws, which were corrected in the papers [PE74], [LR87], [GC03].

3.2 Limitations of the Shannon GPAC

The Shannon GPAC is regarded as an important and powerful method of analog computation, thanks largely to Theorems 15 and 16. Despite this, many authors have pointed out some limitations to the model. For example, the *gamma function*

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

is not differentially algebraic (proven in [Höl86]), and so cannot be generated by a GPAC (as noted by Shannon himself in [Sha41]). However, one could expect that, in a 'sensible' model of computability on continuous data, this function would be computable. Indeed, the gamma function is *effectively computable* in the sense of computable analysis, a branch of analog computability studied by Pour-El, Richards [PER89], Weihrauch [Wei00], Grzegorczyk [Grz55, Grz57], Lacombe [Lac55a, Lac55b, Lac55c], Tucker, Zucker [TZ07], among others. Some authors have addressed this limitation, and Graça [Gra04] showed that the gamma function can indeed be considered as GPAC computable if

> ...we redefine our notion of GPAC-computability in a manner that it matches more closely the philosophy underlying computable analysis...

There is, however, another limitation with the Shannon GPAC, that appears to have been overlooked by Shannon, Pour-El and others. It lies in the fact that the Shannon GPAC can fundamentally

reason only about real-valued functions of one independent variable t. Ironically, it was stated in [Sha41] and [PE74] that the generalization to more than one independent variable only requires an obvious modification, but this is by no means the case. In fact, it is hard to conceive a realistic physical interpretation for a formalism involving two (or more) independent "time" variables.

We briefly remark that this limitation was pointed out in [Rub93]. Rubel says

> For one thing, the GPAC works in one ("time") variable only, while the EAC [Extended Analog Computer] produces functions of any finite number of real variables.

To address this problem, Rubel defined what he called an Extended Analog Computer (EAC).[1] However, Rubel stressed that

> ... the EAC is a *conceptual* computer - the extent to which it can be realized by actual physical, chemical, or biological devices or systems remains to be investigated.

A different attempt to deal with this problem was recently proposed by Bournez, Graça and Pouly [Pou15, BGP16]. In their approach, channels can carry a n-variable real-valued data stream of type $\mathbb{R}^n \to \mathbb{R}$. In this way, they were able to introduce functions with multiple variables by extending the *input space*; for example, replacing $C^1([0,T],\mathbb{R})$ with $C^1([0,X_1]\times\ldots\times[0,X_n],\mathbb{R})$. This seems to be a very natural way of generalizing the Shannon GPAC.

Despite the fact that [BGP16, Definition 14] uses Jacobians (which imply independent variables), we would like to point out that their model can still be re-expressed in terms of only one (implicit "time") variable. This idea is present in [BGP16, Examples 12 and 13]. It also occurs in [BGP16, Remark 15], where it is explained that the value of a generable function y at a given point x can be obtained by solving an initial value problem in one independent variable (this is done by introducing a smooth curve γ from x_0, an initial point, to x).

In this thesis we adopt an approach which in some way is orthogonal to the one in [BGP16]; our idea is to extend the *output space*, that is, replacing $C^1([0,T],\mathbb{R})$ with $C^1([0,T],\mathcal{X})$, where $\mathcal{X}$ is a metric vector space. For example, we can think of $\mathcal{X}$ as the space of continuous real-valued functions on a bounded domain $\Omega\subset\mathbb{R}^n$, that is, $\mathcal{X}=C(\Omega,\mathbb{R})$. In this way, our channels will now carry $\mathcal{X}$-valued streams of data $u:[0,T]\to\mathcal{X}$, which correspond to functions of $n+1$ real variables, under the *uncurrying*

$$[0,T]\to(\Omega\to\mathbb{R})\simeq[0,T]\times\Omega\to\mathbb{R}.$$

It is evident that one of the variables, namely the "time" variable, plays a different role from the others. Our approach is, to some extent, motivated by the theory of partial differential equations, in which some fundamental problems (such as the heat equation, wave equation and Schrödinger equation) can be expressed as time evolution problems in a function space.

3.3 Data channels in function spaces

In this section we present various possibilities of complete metric vector spaces $\mathcal{X}$. We hope that this can provide some intuition on how exactly we intend to generalize Shannon's results to functions of more than one variable. The abundance of possibilities can also be seen as an indication of the broadness of our methods.

Definition 3.3.1 (Domain). A *domain* is an open, connected subset of $\mathbb{R}^d$.

[1] An implementation of the EAC (or at least, of some of its components) has been achieved with the work of Mills, [Mil08].

We will usually denote domains by the Greek letter Ω. In mathematical analysis, there are two common classes of domains, for which different techniques are applicable:

- *unbounded* domains, such as $\Omega = \mathbb{R}^d$;
- *bounded* domains, such as $\Omega = [0,1]^d$;

In this section, we shall restrict our attention to the two examples above, while working in one spatial dimension $d = 1$.

We now present the following types of data spaces: C, C^p and C^∞ spaces, L^2, H^p and H^∞ spaces.

Definition 3.3.2. For a domain $\Omega \subseteq \mathbb{R}$, we denote by $C(\Omega)$ the space of real-valued functions on Ω which are continuous;

- when $\Omega = \mathbb{R}$, this is a Fréchet space with pseudonorms $\|f\|_{C(\mathbb{R}),n} = \sup_{|x|\leq n} |f(x)|$ indexed by $n \in \mathbb{N}$;
- when $\Omega = [0,1]$, this is a Banach space with norm $\|f\|_{C[0,1]} = \sup_{0\leq x\leq 1} |f(x)|$.

Definition 3.3.3. For a domain $\Omega \subseteq \mathbb{R}$ and $k \in \mathbb{N}$, we denote by $C^k(\Omega)$ the space of real-valued functions on Ω which have continuous derivatives of order k;

- when $\Omega = \mathbb{R}$, this is a Fréchet space with pseudonorms $\|f\|_{C^k(\mathbb{R}),j,n} = \sup_{|x|\leq n} |f^{(j)}(x)|$ indexed by $j \in \{0,\ldots,k\}$ and $n \in \mathbb{N}$;
- when $\Omega = [0,1]$, this is a Banach space with norm $\|f\|_{C^k[0,1]} = \sup_{0\leq j\leq k} \sup_{0\leq x\leq 1} |f^{(j)}(x)|$.

Observe that we have $C(\Omega) = C^0(\Omega)$.

Definition 3.3.4. For a domain $\Omega \subseteq \mathbb{R}$, we denote by $C^\infty(\Omega)$ the space of real-valued functions on Ω which have continuous derivatives of any order;

- when $\Omega = \mathbb{R}$, this is a Fréchet space with pseudonorms $\|f\|_{C^\infty(\mathbb{R}),k,n} = \sup_{|x|\leq n} |f^{(k)}(x)|$ indexed by $k \in \mathbb{N}$ and $n \in \mathbb{N}$;
- when $\Omega = [0,1]$, this is a Fréchet space with pseudonorms $\|f\|_{C^\infty[0,1],k} = \sup_{0\leq x\leq 1} |f^{(k)}(x)|$ indexed by $k \in \mathbb{N}$.

Definition 3.3.5. For a domain $\Omega \subseteq \mathbb{R}$, we denote by $L^2(\Omega)$ the space of (equivalence classes under a.e. equality of) real-valued functions on Ω which are square-integrable; whether $\Omega = \mathbb{R}$ or $\Omega = [0,1]$, this is a Banach space with norm $\|f\|^2_{L^2(\Omega)} = \int_\Omega |f(x)|^2 dx$.

Definition 3.3.6. For a domain $\Omega \subseteq \mathbb{R}$, and $k \in \mathbb{N}$, we denote by $H^k(\Omega)$ the space of (equivalence classes under a.e. equality of) real-valued functions on Ω which have square-integrable weak derivatives of order k; whether $\Omega = \mathbb{R}$ or $\Omega = [0,1]$, this is a Banach space with norm $\|f\|^2_{H^k(\Omega)} = \int_\Omega |f(x)|^2 + |f^{(k)}(x)|^2 dx$.

Observe that we have $L^2(\Omega) = H^0(\Omega)$.

Definition 3.3.7. For a domain $\Omega \subseteq \mathbb{R}$, we denote by $H^\infty(\Omega)$ the space of (equivalence classes under a.e. equality of) real-valued functions on Ω which have square-integrable weak derivatives of any order; whether $\Omega = \mathbb{R}$ or $\Omega = [0,1]$, this is a Fréchet space with pseudonorms $\|f\|^2_{H^\infty(\Omega),k} = \int_\Omega |f^{(k)}(x)|^2 dx$ indexed by $k \in \mathbb{N}$.

For the unbounded case $\Omega = \mathbb{R}$, we have the following chains of inclusions:

$$\begin{array}{ccccccccccc} C(\mathbb{R}) & \supset & C^1(\mathbb{R}) & \supset & \cdots & \supset & C^k(\mathbb{R}) & \supset & \cdots & \supset C^\infty(\mathbb{R}). \\ L^2(\mathbb{R}) & \supset & H^1(\mathbb{R}) & \supset & \cdots & \supset & H^k(\mathbb{R}) & \supset & \cdots & \supset H^\infty(\mathbb{R}). \end{array}$$

For the bounded case we have additional inclusions between C^k-spaces and H^k spaces; since bounded functions are integrable in a bounded domain, we have

$$C^k[0,1] \subset H^k[0,1], \text{ for } k \in \mathbb{N};$$

moreover we can apply the Sobolev embedding Theorem [Bré11, Section 9.3]; a version of it (valid for one dimension, and square-integrable functions)[2] states that

$$H^{k+1}[0,1] \subset C^k[0,1], \text{ for } k \in \mathbb{N}.$$

In addition, we may also desire to impose boundary conditions such as Dirichlet, Neumann or periodic conditions. In this section we focus on data that vanishes at the boundary, that is, for $k \in \mathbb{N}$ we consider a restriction to functions g such that $g^{(j)}(0) = g^{(j)}(1) = 0$ for all derivatives up to order k. We use a subscripted zero to indicate we are enforcing boundary vanishing data; thus, we define the spaces $C_0[0,1]$, $C_0^k[0,1]$ for $k \in \mathbb{N}$, $C_0^\infty[0,1]$, $H_0^k[0,1]$, for $k \in \mathbb{N}^+$ and $H_0^\infty[0,1]$.[3]

The Sobolev inclusions also hold for Dirichlet conditions, and in fact we have the relations

$$H_0^{k+1}[0,1] = H^{k+1}[0,1] \cap C_0^k[0,1] \subset C^k[0,1] \cap C_0^k[0,1] = C_0^k[0,1], \text{ for } k \in \mathbb{N}.$$

Thus, we can write the following chains of inclusions:

$$\begin{array}{rcccccccccccc} L^2[0,1] \supset & C[0,1] & \supset & H^1[0,1] & \supset & C^1[0,1] & \supset \cdots \supset & H^k[0,1] & \supset & C^k[0,1] & \supset \cdots \supset & H^\infty[0,1] = C^\infty[0,1] \\ & \cup & & \cup & & \cup & & \cup & & \cup & & \cup \quad \cup \\ C_0[0,1] \supset & H_0^1[0,1] & \supset & C_0^1[0,1] & \supset & \cdots \supset & H_0^k[0,1] & \supset & C_0^k[0,1] & \supset & \cdots \supset & H_0^\infty[0,1] = C_0^\infty[0,1] \end{array}$$

3.4 The $\mathcal{X}$-GPAC

As discussed in Section 3.2 we decide to change our data space $\mathcal{X}$ to become a function space. For most of this chapter we will now focus on a very specific case among those presented in Section 3.3. Namely, we shall take $\mathcal{X}$ to be the space of continuous real functions of a real variable,

$$\mathcal{X} = C(\mathbb{R}).$$

[2] Technically speaking, this inclusion states that any element of H^{k+1}, being an equivalence class of real-valued functions, contains a function which is in C^k.

[3] For Sobolev spaces, we must make another technical remark. From $g \in H^{k+1}[0,1]$, we can use Sobolev inclusions to conclude that g is almost everywhere equal to some function, say $\tilde{g}$, such that $\tilde{g} \in C^k[0,1]$. By $g \in H_0^{k+1}[0,1]$ we simply mean that $\tilde{g}$ satisfies the Dirichlet conditions $\tilde{g}^{(j)}(0) = \tilde{g}^{(j)}(1) = 0$, for all derivatives up to order k. Observe also that $L_0^2[0,1]$ is not a well-defined class, as those 'functions' do not have a well-defined point evaluation.

We make this choice because $C(\mathbb{R})$ is possibly the simplest example from Section 3.3. However, it will still allow us to obtain quite strong equivalence results (Theorem 17 below). For the moment, let us observe that $\mathcal{X}$ is a Fréchet space, with the family of pseudonorms

$$\|g\|_M = \sup_{|x|\leq M} |g(x)|. \tag{3.15}$$

We also consider the subspace $D = C^1(\mathbb{R})$ of continuously differentiable functions. This is also a Fréchet space, with the family of pseudonorms

$$\|g\|_M = \sup_{|x|\leq M} |g(x)| + \sup_{|x|\leq M} |\partial_x g(x)|. \tag{3.16}$$

Note that D is a linear and dense (under the $\mathcal{X}$-pseudonorms) subspace of $\mathcal{X}$.
We can define a differential operator as

$$\begin{array}{rccc} \partial_x : & \mathcal{X} & \rightharpoonup & \mathcal{X} \\ & u(x) & \mapsto & \partial_x u(x). \end{array} \tag{3.17}$$

We establish some important properties of the differential operator:

- ∂_x is a partial function from $\mathcal{X}$ to $\mathcal{X}$, with domain $\mathrm{dom}(\partial_x) = D$;
- ∂_x is linear, that is, for all $u, v \in D$ and $\alpha, \beta \in \mathbb{R}$ we have $\partial_x(\alpha u + \beta v) = \alpha \partial_x u + \beta \partial_x v$;
- ∂_x has dense domain, that is, D is dense in $\mathcal{X}$ (under the topology in $\mathcal{X}$ induced by its pseudonorms);
- ∂_x is a discontinuous operator, as seen in Example 2.1.2;
- ∂_x has a closed graph, that is, if (u_n) is a sequence in D with $u_n \to u$ and $\partial_x u_n \to v$, then $u \in D$ and $\partial_x u = v$;

We include here the definition of closed operator, which will play a role in our notion of well-posedness.

Definition 3.4.1 (Closed operator). Let $\mathcal{X}$ and $\mathcal{Y}$ be metric spaces and $A : \mathcal{X} \rightharpoonup \mathcal{Y}$ a partial function with domain $D(A)$. We say that A is *closed* if its graph is a closed subset of $\mathcal{X} \times \mathcal{Y}$; in other words, if for all sequences (x_n) in $\mathcal{X}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ we have

$$\text{if } \left\{ \begin{array}{ll} x_n \to x & \text{as } n \to \infty, \\ x_n \in D(A) & \text{for all } n, \\ Ax_n \to y & \text{as } n \to \infty; \end{array} \right\} \text{ then } x \in D(A) \text{ and } Ax = y.$$

Hence ∂_x is a closed operator on $\mathcal{X}$. We remind the reader that closedness is, in general, a weaker property than continuity. We also remark that both are equivalent in the space of *total linear* operators between Banach spaces (this basic result is known as the Closed Graph Theorem).

We consider $\mathcal{X}$-streams as elements in the space $C^1([0,T],\mathcal{X})$ of $\mathcal{X}$-valued, continuously differentiable functions, for some $T > 0$. By continuous differentiability we mean that, for $u \in C^1([0,T],\mathcal{X})$, the expression

$$v(t) = \lim_{h\to 0} \frac{u(t+h) - u(t)}{h}$$

is well-defined for all $t \in [0,T]$ and v is a continuous function of time.

In this way, $C^1([0,T],\mathcal{X})$ is a Fréchet space under the family of pseudonorms

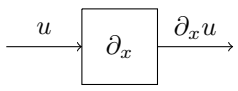
$$\|u\|_{T,M} = \sup_{0\leq t\leq T} \|u(t)\|_M + \sup_{0\leq t\leq T} \|u'(t)\|_M;$$

note that in the right-hand side we consider pseudonorms of $\mathcal{X}$.

We proceed to define the modules for this new setting; at the same time, we introduce a new module based on the differential operator.

Definition 3.4.2 ($\mathcal{X}$**-GPAC modules**)**.** The *$\mathcal{X}$-GPAC modules* are defined as follows.

- for each $g \in \mathcal{X}$, there is a *constant* module (denoted by $\mathbf{cst}_g$ or g) with zero inputs and one output v. We can write $v = \mathbf{cst}_g$, or simply $v = g$; the output is given by
$$v(t) = g;$$
- the *adder* module (denoted by **add** or $+$) has two inputs u, v and one output w. We can write $w = \mathbf{add}(u, v)$, or simply $w = u + v$; the output is given by
$$w(t) = u(t) + v(t);$$
- the *multiplier* module (denoted by **mult** or $\times$) has two inputs u, v and one output w. We can write $w = \mathbf{mult}(u, v)$, or simply $w = u \cdot v$; the output is given by
$$w(t) = u(t)v(t);$$
- the *integrator* module (denoted by **int** or $\int$) has an initial setting g, two inputs u, v and one output w. We can write $w = \mathbf{int}(g, u, v)$ or simply $w = g + \int udv$; the output is given by
$$w(t) = g + \int_0^t u(s)v'(s)ds;$$
- the *differential* module (denoted by **diff** or ∂_x) has one input u and one output v. We can write $v = \mathbf{diff}(u)$ or simply $v = \partial_x u$; the output is given by
$$v(t) = \partial_x u(t).$$



$$\partial_x : C^1([0,T],\mathcal{X}) \rightharpoonup C^1([0,T],\mathcal{X})$$

$$\partial_x(u)(t) = \partial_x u(t)$$

Figure 3.7: The differential module.

We remark that the differential module $\partial_x : C^1([0,T],\mathcal{X}) \rightharpoonup C^1([0,T],\mathcal{X})$ is partially defined and its domain is $C^1([0,T],D)$. As mentioned above, ∂_x is a closed but discontinuous operator.

Remark 3.4.3. The choices adopted here for basic modules in an $\mathcal{X}$-GPAC are slightly different from those originally presented by Shannon (Definition 3.1.1). However, both constructions can be seen to be equivalent. Later on (in Sections 3.7 and 3.8) we shall show how to convert between different formulations of the basic modules.

With the above considerations in mind we can arrive at the desired generalization of the Shannon GPAC. The following Definitions 3.4.4, 3.4.5, 3.4.7 and 3.4.11 should be compared with Definitions 3.1.5, 3.1.8, 3.1.11 and 3.1.12.

Definition 3.4.4 ($\mathcal{X}$**-GPAC**)**.** An $\mathcal{X}$*-valued general purpose analog computer* ($\mathcal{X}$-GPAC) is a network built with the five $\mathcal{X}$-modules (constants, adders, multipliers, integrators and differentials) and $\mathcal{X}$-channels connecting their inputs and outputs, with the following restrictions:

- the only connections allowed are between an output and an input;
- each input may be connected to either zero or one output;

Definition 3.4.5 ($\mathcal{X}$**-GPAC induced operator**)**.** Let $\mathcal{G}$ be an $\mathcal{X}$-GPAC;

- the *constant space* of $\mathcal{G}$ is the cartesian product of the spaces associated with all the initial settings occuring in any integrator. The constant space can be written as $\mathcal{C} = \mathcal{X}^p$, for some $p \geq 0$;
- the *proper input space* of $\mathcal{G}$ is the cartesian product of the spaces associated with all the unconnected input channels. The input space can be written as $\mathcal{I} = C^1([0,T],\mathcal{X})^q$, for some $q \geq 0$;
- the *proper output space* of $\mathcal{G}$ is the cartesian product of the spaces associated with all the unconnected output channels. The output space can be written as $\mathcal{O} = C^1([0,T],\mathcal{X})^m$, for some $m \geq 0$;
- the *mixed space* of $\mathcal{G}$ is the cartesian product of the spaces associated with all the channels which connect some input with some output. The mixed space can be written as $\mathcal{M} = C^1([0,T],\mathcal{X})^r$, for some $r \geq 0$;
- the *induced operator* of $\mathcal{G}$ is the function

$$\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \mathcal{M} \times \mathcal{O}, \quad \Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\tilde{\mathbf{u}}^M, \mathbf{u}^O), \tag{3.18}$$

  where $\mathbf{g}$ is an $\mathcal{X}$-valued vector and each $\mathbf{u}$ is a vector of $\mathcal{X}$-channels. Moreover, each of the components in the codomain of Φ is given by the module with which it is associated. In other words, for each u_i component of either $\tilde{\mathbf{u}}^M$ or $\mathbf{u}^O$,

  - if u_i is associated with the output channel of a constant, then $u_i = g$, where g is the constant associated with that module;
  - if u_i is associated with the output channel of an adder, then $u_i = +(v_1, v_2)$, where v_1 and v_2 are the components in $\mathbf{u}^I$ or $\mathbf{u}^M$ associated with the input channels of that module;
  - if u_i is associated with the output channel of a multiplier, then $u_i = \times(v_1, v_2)$, where v_1 and v_2 are the components in $\mathbf{u}^I$ or $\mathbf{u}^M$ associated with the input channels of that module;
  - if u_i is associated with the output channel of an integrator, then $u_i = \int(g, v_1, v_2)$, where v_1, v_2 are the components in $\mathbf{u}^I$ or $\mathbf{u}^M$ associated with the input channels of that module, and g is the component in $\mathbf{g}$ associated with the initial setting of that module;
  - if u_i is associated with the output channel of a differential, then $u_i = \partial_x(v)$, where v is the component in $\mathbf{u}^I$ or $\mathbf{u}^M$ associated with the input channel of that module.

We remark that Φ may be partially defined; for each differential module occuring in $\mathcal{G}$, there is a component of $\mathcal{I} \times \mathcal{M}$ which is restricted to $C^1([0,T], D)$. Hence, to describe the domain of Φ we must take into special consideration those channels which are inputs of differential modules.

Example 3.4.6. To achieve a better understanding of Definition 3.4.5 we provide an example in Figure 3.8 of an $\mathcal{X}$-GPAC with one constant and four $\mathcal{X}$-channels.

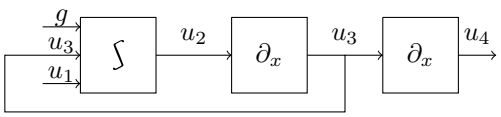


Figure 3.8: An $\mathcal{X}$-GPAC implementing a transport equation, and one spatial derivative of its solution.

In this example there is one constant (associated with the integrator module), one input channel (labeled u_1), two mixed channels (labeled u_2 and u_3) and one output channel (labeled u_4). The induced operator simply formalizes the input/output relation between these channels,

$$\mathcal{C} = \mathcal{X}, \quad \mathcal{I} = C^1(\mathbb{T}, \mathcal{X}), \quad \mathcal{M} = C^1(\mathbb{T}, \mathcal{X})^2, \quad \mathcal{O} = C^1(\mathbb{T}, \mathcal{X});$$
$$\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \mathcal{M} \times \mathcal{O}$$
$$\Phi(g, u_1, u_2, u_3) = \left(g + \int u_3 du_1, \partial_x u_2, \partial_x u_3 \right) = (\tilde{u_2}, \tilde{u_3}, u_4).$$

Definition 3.4.4 gives a lot of freedom in the construction of $\mathcal{X}$-GPACs and it turns out that not all of the possible networks lead to 'valid and interesting' $\mathcal{X}$-GPACs (similarly to the fact that not all ASCII expressions lead to 'valid and interesting' computer programs). Thus we present a well-posedness-like notion to restrict the space of $\mathcal{X}$-GPACs that we wish to consider.

Definition 3.4.7 (Quasi-well-posedness of $\mathcal{X}$-GPAC). Let $\mathcal{G}$ be an $\mathcal{X}$-GPAC and $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \mathcal{M} \times \mathcal{O}$ be its induced operator. Let $U \subseteq \mathcal{C} \times \mathcal{I}$. We say that $\mathcal{G}$ is *quasi-well-posed* on U if

- (*existence*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, there exists $(\mathbf{u}^M, \mathbf{u}^O) \in \mathcal{M} \times \mathcal{O}$ such that
$$\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\mathbf{u}^M, \mathbf{u}^O); \tag{3.19}$$
- (*uniqueness*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, the tuple $(\mathbf{u}^M, \mathbf{u}^O)$ such that (3.19) holds is unique;
- (*closedness*) the map $(\mathbf{g}, \mathbf{u}^I) \mapsto (\mathbf{u}^M, \mathbf{u}^O)$, with domain U and codomain $\mathcal{M} \times \mathcal{O}$, given as the unique solution of (3.19), defines a closed operator.

We may refer to (3.19) as the *fixed point equation*; note that the mixed variables $\mathbf{u}^M$ are the only ones that appear in both sides of the equation.

If, as in Definition 3.1.11, we required continuity instead of closedness (that is, if we required U to be an open set, and the map $(\mathbf{g}, \mathbf{u}^I) \mapsto (\mathbf{u}^M, \mathbf{u}^O)$ to be continuous), then our definition would match the three usual principles for *well-posedness* - existence, uniqueness, continuity of solutions - as presented by Hadamard [Had52] (see also [CH53]). Instead, we choose to use closedness (and a non-necessarily open domain U), which is a strictly weaker criterion, hence the term 'quasi-well-posed'. The reason for choosing closedness is the presence of the differential module, which defines a closed function of type $\mathcal{X} \rightharpoonup \mathcal{X}$ which is not continuous.

Admittedly, some could argue that a discontinuous operation is of little interest in the study of computable functions on continuous spaces. We must then make the important remark that continuity can be obtained from closedness (and thus well-posedness can be obtained from quasi-well-posedness) if we define a finer topology in the domain, induced by graph (pseudo)norms. To be precise, we recall the following basic result in functional analysis.

Proposition 3.4.8 ([RR06, p. 240, Exercise 8.7])**.** *Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ be Banach spaces and $A : \mathcal{X} \rightharpoonup \mathcal{Y}$ a closed linear operator with domain $D(A)$. Then*

- *$(D(A), \|\cdot\|_{G(A)})$ is a Banach space with the graph norm given by*

$$\|x\|_{G(A)} = \|x\|_{\mathcal{X}} + \|Ax\|_{\mathcal{Y}};$$

- *the restriction $A : D(A) \to \mathcal{Y}$ is a continuous linear map between Banach spaces.*

Proof. We first prove that $\|\cdot\|_{G(A)}$ defines a norm:

- if $x \in D(A)$, then $\|x\|_{G(A)} = 0$ iff $\|x\|_{\mathcal{X}} + \|Ax\|_{\mathcal{Y}} = 0$ iff $\|x\|_{\mathcal{X}} = 0$ iff $x = 0$;
- if $x \in D(A)$ and $\alpha \in \mathbb{R}$, then $\|\alpha x\|_{G(A)} = \|\alpha x\|_{\mathcal{X}} + \|A(\alpha x)\|_{\mathcal{Y}} = |\alpha|\|x\|_{\mathcal{X}} + |\alpha|\|Ax\|_{\mathcal{Y}} = |\alpha|\|x\|_{G(A)}$;
- if $x, y \in D(A)$, then $\|x+y\|_{G(A)} = \|x+y\|_{\mathcal{X}} + \|A(x+y)\|_{\mathcal{Y}} \leq \|x\|_{\mathcal{X}} + \|y\|_{\mathcal{X}} + \|Ax\|_{\mathcal{Y}} + \|Ay\|_{\mathcal{Y}} = \|x\|_{G(A)} + \|y\|_{G(A)}$.

Next we prove that $D(A)$ is complete. Let $\{x_n\}_n$ be a Cauchy sequence in $D(A)$, so that $\|x_{N+k} - x_N\|_{G(A)} = \|x_{N+k} - x_N\|_{\mathcal{X}} + \|A(x_{N+k} - x_N)\|_{\mathcal{Y}}$ vanishes (uniformly on k) as $N \to \infty$. In particular, we must have that $\{x_n\}_n$ is a Cauchy sequence in $\mathcal{X}$ (under the $\mathcal{X}$-norm) and $\{Ax_n\}$ is a Cauchy sequence in $\mathcal{Y}$. Since $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces it follows that there exist $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $x_n \to x$ and $Ax_n \to y$. By closedness of A we conclude that $x \in D(A)$ and $Ax = y$, so that $x_n \to x$ in $D(A)$ (under the graph norm). Thus $(D(A), \|\cdot\|_{G(A)})$ is a Banach space.

Finally we prove that $A : D(A) \to \mathcal{Y}$ is continuous. Let $\{x_n\}_n$ be a sequence in $D(A)$ and $x \in D(A)$ such that $x_n \to x$ in $D(A)$. Then $\|x_n - x\|_{G(A)} \to 0$, which implies $\|x_n - x\|_{\mathcal{X}} \to 0$ and $\|A(x_n - x)\|_{\mathcal{Y}} \to 0$, so that, in particular, $Ax_n \to Ax$ in $\mathcal{Y}$. □

Proposition 3.4.9. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Fréchet spaces with families of pseudonorms $\{\|\cdot\|_{\mathcal{X},n}\}_n$ and $\{\|\cdot\|_{\mathcal{Y},m}\}_m$ and $A : \mathcal{X} \rightharpoonup \mathcal{Y}$ a closed linear operator with domain $D(A)$. Then*

- *$D(A)$ is a Fréchet space with graph pseudonorms given by*

$$\|x\|_{G(A),n,m} = \|x\|_{\mathcal{X},n} + \|Ax\|_{\mathcal{Y},m};$$

- *the restriction $A : D(A) \to \mathcal{Y}$ is a continuous linear map between Fréchet spaces.*

Proof. The proof is similar to that of Proposition 3.4.8. □

Proposition 3.4.10. *Let $\mathcal{X}$ and $\mathcal{Y}$ be complete metric spaces with metrics $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ and $A : \mathcal{X} \rightharpoonup \mathcal{Y}$ a closed operator with domain $D(A)$. Then*

- *$D(A)$ is a complete metric space with graph metric given by*

$$d_{G(A)}(x_1, x_2) = d_{\mathcal{X}}(x_1, x_2) + d_{\mathcal{Y}}(Ax_1, Ax_2);$$

- *the restriction $A : D(A) \to \mathcal{Y}$ is a continuous linear map between complete metric spaces.*

Proof. The proof is similar to that of Proposition 3.4.8. □

This finer topology is usually equivalent to a topology of interest in the domain space. For example, consider the differential operator $\partial_x : C(\mathbb{R}) \rightharpoonup C(\mathbb{R})$ given by (3.17), which is closed but not continuous under the usual family of pseudonorms in $C(\mathbb{R})$. However, it becomes a continuous operator if we restrict it to the space $C^1(\mathbb{R})$ and consider the graph pseudonorms

$$\|f\|_{n,m} = \|f\|_n + \|\partial_x f\|_m = \sup_{|x|\leq n} |f(x)| + \sup_{|x|\leq m} |\partial_x f(x)|,$$

under whose topology $C^1(\mathbb{R})$ is a Fréchet space. Moreover, this family of pseudonorms can be seen to be equivalent to the usual family of pseudonorms in $C^1(\mathbb{R})$ given by (3.16).

We shall then adopt the notion of quasi-well-posedness in this chapter, while reminding ourselves that, if needed, we can in principle express our results in terms of well-posed operators. Later on, in Chapter 5, we will consider a slightly different notion of induced operator for which well-posedness is regained.

The final step in this section is to assign semantics to $\mathcal{X}$-GPACs, that is, to define the notion of $\mathcal{X}$-GPAC-generable functions.

Definition 3.4.11 (**Semantics of $\mathcal{X}$-GPAC**)**.**

(a) Let $\mathcal{G}$ be an $\mathcal{X}$-GPAC and $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \mathcal{M} \times \mathcal{O}$ be its induced operator. Let $U \subseteq \mathcal{C} \times \mathcal{I}$ such that $\mathcal{G}$ is quasi-well-posed on U. We define the *specification* of $\mathcal{G}$ as the (partial) function

$$\begin{aligned} &F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}; \\ &F(\mathbf{g}, \mathbf{u}^I) = (\mathbf{u}^M, \mathbf{u}^O), \end{aligned}$$

whose domain is U and where $(\mathbf{u}^M, \mathbf{u}^O)$ is given by (3.19). We also say that $\mathcal{G}$ *generates* F on $U = \mathrm{dom}(F)$.

(b) A function $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$ is *$\mathcal{X}$-GPAC-generable* if there exists an $\mathcal{X}$-GPAC $\mathcal{G}$ such that $\mathcal{G}$ is quasi-well-posed on the domain of F and F is the specification of $\mathcal{G}$.

We remark that every integrator in an $\mathcal{X}$-GPAC has an initial setting (which is one of the constants in the space $\mathcal{C}$) and an output (which is one of the mixed/output channels in $\mathcal{M} \times \mathcal{O}$). Since we can define an injective map from the initial settings to the mixed/output channels, we have the following basic property.

Proposition 3.4.12. *If $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$ is $\mathcal{X}$-GPAC-generable, then* $\dim(\mathcal{C}) \leq \dim(\mathcal{M}) + \dim(\mathcal{O})$.

3.5 Normal form systems

In the previous section we have defined the class of $\mathcal{X}$-GPAC-generable functions, which constitute our space of interest. We can state our objective as follows.

Problem. Characterize the class of $\mathcal{X}$-GPAC-generable functions in terms of a suitable generalization of the class of differential algebraic equations.

In the study of the GPAC, an intermediate step is usually taken in the transition from generable functions to differential algebraic equations. For example, in [Sha41] a system of equations called

a *fundamental solvability condition* was considered (Theorem I in that paper). Also in [PE74] a similar system of equations is used in the actual definition of GPAC generable functions (Definition 10 in that paper; see also Lemma 3.1.17 and Remark 3.1.19) instead of the usual definition involving analog networks. We shall generalize that notion into our framework, and refer to the resulting objects as normal form systems.

Definition 3.5.1 (**Normal form equation**)**.** Let $N \in \mathbb{N}$. A *normal form equation* on the N variables $y_1, \dots, y_N$ is an equation of the form

$$\sum_{i=0}^{N} \sum_{j=1}^{N} b_{ij} y_i y_j' + \sum_{j=1}^{N} c_j \partial_x y_j' = 0, \tag{3.20}$$

under the conventions that $y_0 \equiv 1$ and b_{ij}, c_j are real numbers.

We remark that $y_0 \equiv 1$ is not a variable, just a notational convenience to obtain a more compact equation. We also note that index i starts at 0 whereas index j starts at 1; thus (3.20) will not include the term y_0' (which would be equal to 0, and therefore irrelevant).

We also alert the reader to our choice of terms of the form $\partial_x y_j'$ in the second summation in (3.20), with derivatives both in time and space. One could think that a similar definition of normal form equations with terms of the form $\partial_x y_j$ would be more 'natural'. In fact, both definitions would be equivalent, and we could move from the former to the latter by adding extra variables (namely, one would have to add the extra variable $\mathbf{t}$ that specifies "linear time", $\mathbf{t}(t, x) = t$). The main reason for choosing (3.20) as our template for normal form equations is practicality, as it makes the proof of Lemma 3.5.7 (below) much simpler.

Here are some examples of normal form equations:

$$\begin{aligned} y_1' &= y_1 y_2'; \\ y_1' &= y_2' + y_3'; \\ y_1' &= \partial_x y_1'. \end{aligned}$$

Definition 3.5.2 (**Normal form system over $\mathcal{X}$**)**.** Let $K, L \in \mathbb{N}$ and $N = K + L$. A *normal form system* (NFS) on the N variables $y_1, \dots, y_N$ is a system of the form

$$\begin{cases} \displaystyle\sum_{i=0}^{N} \sum_{j=1}^{N} b_{ij\ell} y_i y_j' + \sum_{j=1}^{N} c_{j\ell} \partial_x y_j' = 0 & \text{, for } \ell = 1, \dots, L; \\ y_{K+\ell}(0) = g_\ell & \text{, for } \ell = 1, \dots, L, \end{cases} \tag{3.21}$$

under the conventions that $y_0 \equiv 1$, $b_{ij\ell}$, $c_{j\ell}$ are real numbers and $g_\ell \in \mathcal{X}$.

We say that $g_1, \dots, g_L$ are the *(initial) constants*, $y_1, \dots, y_K$ are the *inputs* or *independent variables* and $y_{K+1} \dots, y_{K+L}$ are the *outputs* or *dependent variables*.

We briefly remark that, in a well-posed system of equations, the number of equations and the number of unknowns must be the same. In the previous definition, these correspond to the outputs $y_{K+1}, \dots, y_{K+L}$; hence there must be L equations.

Definition 3.5.3 (**Quasi-well-posedness of NFS**)**.** Let $K, L \in \mathbb{N}$ and $N = K + L$. Let $\mathcal{N}$ be an NFS given by (3.21) with K inputs and L outputs and consider the spaces

$$\mathcal{C} = \mathcal{X}^L, \quad \mathcal{I} = C^1([0, T], \mathcal{X})^K, \quad \mathcal{O} = C^1([0, T], \mathcal{X})^L.$$

Let $U \subseteq \mathcal{C} \times \mathcal{I}$. We say that $\mathcal{N}$ is *quasi-well-posed* on U if

- (*existence*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, there exists $\mathbf{u}^O \in \mathcal{O}$ such that (3.21) holds for $(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^O)$, where $\mathbf{g} = (g_1, \ldots, g_L)$, $\mathbf{u}^I = (y_1, \ldots, y_K)$, $\mathbf{u}^O = (y_{K+1}, \ldots, y_{K+L})$;
- (*uniqueness*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, the tuple $\mathbf{u}^O$ such that (3.21) holds is unique;
- (*closedness*) the map $(\mathbf{g}, \mathbf{u}^I) \mapsto \mathbf{u}^O$, with domain U and codomain $\mathcal{O}$, given as the unique solution of (3.21), defines a closed operator.

Definition 3.5.4 (**Semantics of NFS**).

(a) Let $K, L \in \mathbb{N}$ and $N = K + L$. Let $\mathcal{N}$ be an NFS with K inputs and L outputs. Let $U \subseteq \mathcal{C} \times \mathcal{I}$ such that $\mathcal{N}$ is quasi-well-posed on U. We define the *solution* of $\mathcal{N}$ as the (partial) function

$$F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O};$$
$$F(\mathbf{g}, \mathbf{u}^I) = \mathbf{u}^O,$$

whose domain is U and where $\mathbf{u}^O$ is given by (3.21). We also say that $\mathcal{N}$ *generates* F on $U = \mathrm{dom}(F)$.

(b) A function $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O}$ is *NFS-generable* if there exists an NFS $\mathcal{N}$ such that $\mathcal{N}$ is quasi-well-posed on the domain of F and F is the solution of $\mathcal{N}$.

We remark that in an NFS, there is a bijection between the initial conditions and the outputs; thus we have the following basic property (cf. Proposition 3.4.12).

Proposition 3.5.5. *If $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O}$ is NFS-generable, then* $\dim(\mathcal{C}) = \dim(\mathcal{O})$.

We have established semantics for NFS, and the next step is to show that every $\mathcal{X}$-GPAC-generable function is a projection of an NFS-generable function, as in the following definition.

Definition 3.5.6 (**Function projection**). Let $F : A \rightharpoonup B$ and $F' : A \times A' \rightharpoonup B \times B'$, and let

$$\begin{array}{rcccl} G(F) & = & \{(a,b) : a \in \mathrm{dom}(F) \text{ and } F(a) = b\} & \subseteq & A \times B, \\ G(F') & = & \{(a,a',b,b') : (a,a') \in \mathrm{dom}(F') \text{ and } F'(a,a') = (b,b')\} & \subseteq & A \times A' \times B \times B' \end{array}$$

be their graphs. We say that F is a *projection* of F' if for all $(a, b) \in A \times B$,

- if $(a, b) \notin G(F)$ then for all $a' \in A'$, $b' \in B'$ we have $(a, a', b, b') \notin G(F')$;
- if $(a, b) \in G(F)$ then there exist unique $a' \in A'$ and $b' \in B'$ such that $(a, a', b, b') \in G(F')$.

We briefly remark that the notion of projection induces a partial order in the class of functions. We have the following lemma.

Lemma 3.5.7 (**$\mathcal{X}$-GPAC-generable implies NFS-generable**). *Let $F_G : \mathcal{C}_G \times \mathcal{I}_G \rightharpoonup \mathcal{M}_G \times \mathcal{O}_G$ be $\mathcal{X}$-GPAC-generable with domain $U_G \subseteq \mathcal{C}_G \times \mathcal{I}_G$. Then there exists $F_N : \mathcal{C}_N \times \mathcal{I}_N \rightharpoonup \mathcal{O}_N$ which is NFS-generable with domain $U_N \subseteq \mathcal{C}_N \times \mathcal{I}_N$ with the following properties:*

- $\dim(\mathcal{I}_N) = \dim(\mathcal{I}_G)$ *and* $\dim(\mathcal{O}_N) = \dim(\mathcal{M}_G) + \dim(\mathcal{O}_G)$*;*
- *for every $(\mathbf{g}, \mathbf{u}^I) \in \mathcal{C}_G \times \mathcal{I}_G$ such that $(\mathbf{g}, \mathbf{u}^I) \notin U_G$, we have $(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \notin U_N$ for any $\mathcal{X}$-vector $\mathbf{g}^*$;*
- *for every $(\mathbf{g}, \mathbf{u}^I) \in \mathcal{C}_G \times \mathcal{I}_G$ such that $(\mathbf{g}, \mathbf{u}^I) \in U_G$, there exists a unique $\mathcal{X}$-vector $\mathbf{g}^*$ such that $(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \in U_N$; moreover, we have*

$$F_G(\mathbf{g}, \mathbf{u}^I) = F_N(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I).$$

In other words, F_G is a projection of F_N.

Proof. Let $F_G : \mathcal{C}_G \times \mathcal{I}_G \rightharpoonup \mathcal{M}_G \times \mathcal{O}_G$ be $\mathcal{X}$-GPAC-generable with domain $U_G \subseteq \mathcal{C}_G \times \mathcal{I}_G$. Denote by $\mathcal{G}$ the $\mathcal{X}$-GPAC that generates F_G and $\Phi : \mathcal{C}_G \times \mathcal{I}_G \times \mathcal{M}_G \rightharpoonup \mathcal{M}_G \times \mathcal{O}_G$ the induced operator.

The important idea of the proof is understanding how to write the equational specification of Φ as an NFS. We apply the following conversions, for every u appearing in $\mathcal{M}_G \times \mathcal{O}_G$:

$$
\begin{aligned}
u = g \quad &\rightsquigarrow \quad \begin{cases} u' = 0 \\ u(0) = g \end{cases} \\
u = v + w \quad &\rightsquigarrow \quad \begin{cases} u' = v' + w' \\ u(0) = v(0) + w(0) \end{cases} \\
u = v \cdot w \quad &\rightsquigarrow \quad \begin{cases} u' = vw' + wv' \\ u(0) = v(0) \cdot w(0) \end{cases} \\
u = g + \int v dw \quad &\rightsquigarrow \quad \begin{cases} u' = vw' \\ u(0) = g \end{cases} \\
u = \partial_x v \quad &\rightsquigarrow \quad \begin{cases} u' = \partial_x v' \\ u(0) = \partial_x v(0) \end{cases}
\end{aligned}
$$

We observe that the input/output relation of each module of $\mathcal{G}$ can be written as a normal form equation coupled with an initial condition. Therefore, we can construct an NFS, $\mathcal{N}$, from $\mathcal{G}$, including an initial condition for each channel.

However, the constant space $\mathcal{C}_G$ appearing in the specification of $\mathcal{G}$ only takes into account those constants appearing as initial settings of integrators. In order to define the solution mapping, Φ_N, we need to extend the constant space to include the initial settings of the other types of operations, as they appear in the list of conversions above.

Let $\mathcal{I}_N = C^1([0,T], \mathcal{X})^K$ and $\mathcal{O}_N = C^1([0,T], \mathcal{X})^L$, where K, L are the number of inputs and outputs of $\mathcal{N}$. By construction each input of $\mathcal{N}$ corresponds to an input channel of $\mathcal{G}$ and each output of $\mathcal{N}$ corresponds to either a mixed or output channel of $\mathcal{G}$. Therefore, we have that $\mathcal{I}_N = \mathcal{I}_G$ and $\mathcal{O}_N = \mathcal{M}_G \times \mathcal{O}_G$, so that $\dim(\mathcal{I}_N) = \dim(\mathcal{I}_G)$ and $\dim(\mathcal{O}_N) = \dim(\mathcal{M}_G) + \dim(\mathcal{O}_G)$, which proves the first bullet.

The next step is to find a suitable U_N for the domain of F_N. By quasi-well-posedness of $\mathcal{G}$ on U_G, we know that, for every $(\mathbf{g}, \mathbf{u}^I) \in U_G$, there exists a unique $(\mathbf{u}^M, \mathbf{u}^O) \in \mathcal{M}_G \times \mathcal{O}_G$ such that $\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\mathbf{u}^M, \mathbf{u}^O)$. Thus, if we define $(\mathbf{g}, \mathbf{g}^*)$ as the vector of initial conditions[4] of $(\mathbf{u}^M, \mathbf{u}^O)$, we can infer that $\mathbf{g}^*$ depends uniquely on $(\mathbf{u}^M, \mathbf{u}^O)$, and thus it depends uniquely on $(\mathbf{g}, \mathbf{u}^I)$. With this in mind we define

$$
\begin{aligned}
U_N &= \{(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) : (\mathbf{g}, \mathbf{u}^I) \in U_G \text{ and } (\mathbf{g}, \mathbf{g}^*) = F_G(\mathbf{g}, \mathbf{u}^I) \restriction_{t=0}\}; \\
F_N(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) &= \begin{cases} F_G(\mathbf{g}, \mathbf{u}^I) & \text{if } (\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \in U_N; \\ \text{undefined} & \text{otherwise.} \end{cases}
\end{aligned}
$$

By the above construction, U_N and F_N satisfy the second and third bullets, and all is left is to show that F_N is the solution of $\mathcal{N}$ on U_N. By quasi-well-posedness of $\mathcal{G}$ on U_G, it is clear that for $(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \in U_N$, the tuple $(\mathbf{u}^M, \mathbf{u}^O) \in \mathcal{O}_N$ that solves the NFS exists, is unique and given by $F_G(\mathbf{g}, \mathbf{u}^I)$.

To prove closedness of F_N, consider a sequence $(\mathbf{g}_n, \mathbf{g}_n^*, \mathbf{u}_n^I) \in U_N$ such that

[4] Possibly after reordering the mixed and output channels; this can be done without loss of generality.

$$(\mathbf{g}_n, \mathbf{g}_n^*, \mathbf{u}_n^I) \to (\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \text{ and } F_N(\mathbf{g}_n, \mathbf{g}_n^*, \mathbf{u}_n^I) \to (\mathbf{u}^M, \mathbf{u}^O);$$

then we have

$$(\mathbf{g}_n, \mathbf{u}_n^I) \to (\mathbf{g}, \mathbf{u}^I) \text{ and } F_G(\mathbf{g}_n, \mathbf{u}_n^I) = F_N(\mathbf{g}_n, \mathbf{g}_n^*, \mathbf{u}_n^I) \to (\mathbf{u}^M, \mathbf{u}^O).$$

By closedness of F_G, we have

$$(\mathbf{g}, \mathbf{u}^I) \in U_N \text{ and } F_G(\mathbf{g}, \mathbf{u}^I) = (\mathbf{u}^M, \mathbf{u}^O);$$

Now define $(\mathbf{u}_n^M, \mathbf{u}_n^O) = F_G(\mathbf{g}_n, \mathbf{u}_n^I)$, so that

$$(\mathbf{g}_n, \mathbf{g}_n^*) = F_G(\mathbf{g}_n, \mathbf{u}_n^I) \restriction_{t=0} = (\mathbf{u}_n^M, \mathbf{u}_n^O) \restriction_{t=0};$$

by taking limits, we conclude that $(\mathbf{g}, \mathbf{g}^*) = (\mathbf{u}^M, \mathbf{u}^O) \restriction_{t=0}$. Thus,

$$(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) \in U_N \text{ and } F_N(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I) = (\mathbf{u}^M, \mathbf{u}^O),$$

which concludes the proof. □

Example 3.5.8. In order to better understand the construction in the proof of Lemma 3.5.7, we apply it to the $\mathcal{X}$-GPAC seen on Example 3.4.6 (repeated in Figure 3.9).

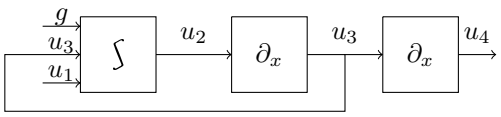


Figure 3.9: An $\mathcal{X}$-GPAC implementing a transport equation, and one spatial derivative of its solution.

It can be checked that this $\mathcal{X}$-GPAC is well-posed for $u_1 \in C^1([0,T], \mathcal{X})$ and $g \in C^2(\mathbb{R})$. It generates the function $F_G : \mathcal{X}^1 \times C^1([0,T], \mathcal{X})^1 \rightharpoonup C^1([0,T], \mathcal{X})^3$, given by $F_G(g, u_1) = (u_2, u_3, u_4)$, where

$$u_2(t,x) = g(x + u_1(t) - u_1(0)), \quad u_3(t,x) = g'(x + u_1(t) - u_1(0)), \quad u_4(t,x) = g''(x + u_1(t) - u_1(0));$$

if u_1 is linear time, $u_1 = \mathbf{t}$, this has the simpler form

$$u_2(t,x) = g(x+t), \quad u_3(t,x) = g'(x+t), \quad u_4(t,x) = g''(x+t).$$

Let us construct an NFS with solution F_N such that F_G is a projection of F_N. The $\mathcal{X}$-GPAC generates the equational relations

$$u_2 = g + \int_0^t u_3 du_1, \quad u_3 = \partial_x u_2, \quad u_4 = \partial_x u_3; \tag{3.22}$$

which can be converted into an NFS with three constants, one input and three outputs,

$$\begin{array}{lll} u_2' = u_3 u_1', & u_3' = \partial_x u_2', & u_4' = \partial_x u_3', \\ u_2(0) = g, & u_3(0) = g_3, & u_4(0) = g_4. \end{array} \tag{3.23}$$

We can see that the constant space of the NFS has three parameters g, g_3 and g_4, which is two more than in the constant space of the $\mathcal{X}$-GPAC. Therefore any solution of the NFS must be of type $\mathcal{X}^3 \times C^1([0,T],\mathcal{X})^1 \rightharpoonup C^1([0,T],\mathcal{X})^3$.

We also see that if (g, u_1, u_2, u_3, u_4) produces a specification of the $\mathcal{X}$-GPAC, then (u_1, u_2, u_3, u_4) produces a solution of the NFS for the initial conditions g, $g_3 = u_3(0)$, $g_4 = u_4(0)$; in other words, the NFS generates a function F_N such that

$$\text{if } F_G(g, u_1) = (u_2, u_3, u_4), \text{ then } F_N(g, u_3(0), u_4(0), u_1) = (u_2, u_3, u_4).$$

Thus F_G is a projection of F_N, as expected.

3.6 Partial differential algebraic equations

In this section we will define partial differential algebraic systems, which will prove to be the correct generalization of differentially algebraic equations for our purposes, as will be made clear from our main result (Theorem 17).

Definition 3.6.1 (**Partial differential algebraic equation**). Let $N \in \mathbb{N}$. A *partial differential algebraic equation* (PDAE) on the N variables $y_1, \dots, y_N$ is an equation of the form

$$P(t, y_1, \dots, y_N, \dots, \partial_x^{\alpha_1} y_1^{(\beta_1)}, \dots, \partial_x^{\alpha_N} y_N^{(\beta_N)}) = 0, \tag{3.24}$$

where P is a polynomial in $y_1, \dots, y_N$ and some of their derivatives, with real coefficients.

Definition 3.6.2 (**System of PDAEs**). Let $K, L \in \mathbb{N}$ and $N = K + L$. A *partial differential algebraic system* (PDAS), also referred to as *system of PDAEs*, on the N variables $y_1, \dots, y_K, z_1, \dots, z_L$ is a system of the form

$$\begin{cases} P_\ell(t, y_1, \dots, y_K, z_1, \dots, z_L \dots, \partial_x^{\alpha_1} y_1^{(\beta_1)}, \dots, \partial_x^{\alpha_N} z_L^{(\beta_N)}) = 0 & \text{, for } 1 \le \ell \le L; \\ z_\ell^{(\beta)}(0) = g_{\ell,\beta} & \text{, for } 1 \le \ell \le L \text{ and } 0 \le \beta < \beta_\ell, \end{cases} \tag{3.25}$$

under the conventions that P_ℓ are polynomials in $y_1, \dots, y_K, z_1, \dots, z_L$ and some of their derivatives, with real coefficients, and $g_{\ell,\beta} \in \mathcal{X}$.

We say that $y_1, \dots, y_K$ are the *inputs* or *independent variables* and $z_1 \dots, z_L$ are the *outputs* or *dependent variables*.

We provide a short explanation on the notation in the previous definition. Each variable y_k can appear in the polynomial expressions with space derivatives of order at most α_k and time derivatives of order at most β_k, and similarly for z_ℓ; for each output z_ℓ, we need to provide β_ℓ initial conditions; they correspond to the values of z_ℓ and its time derivatives $z_\ell^{(\beta)}$ of order up to $\beta_\ell - 1$ at time $t = 0$,

$$z_\ell(0), \quad z_\ell'(0), \quad \dots, \quad z_\ell^{(\beta_\ell - 1)}(0).$$

This is a standard assumption in the theory of PDEs and is a necessary condition for well-posedness (with fewer initial conditions, the system is underdetermined; with more initial conditions, the system is overdetermined).

Definition 3.6.3 (**Quasi-well-posedness of PDAS**). Let $K, L \in \mathbb{N}$ and $N = L + K$. Let $\mathcal{P}$ be a PDAS given by (3.25) with K inputs, L outputs and J initial conditions and consider the spaces

$$\mathcal{C} = \mathcal{X}^J, \quad \mathcal{I} = C^1([0,T],\mathcal{X})^K, \quad \mathcal{O} = C^1([0,T],\mathcal{X})^L.$$

Let $U \subseteq \mathcal{C} \times \mathcal{I}$. We say that $\mathcal{P}$ is *quasi-well-posed* on U if

- (*existence*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, there exists $\mathbf{u}^O \in \mathcal{O}$ such that (3.25) holds for $(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^O)$, where $\mathbf{g}$ is the vector of initial conditions, $\mathbf{u}^I = (y_1, \ldots, y_K)$, $\mathbf{u}^O = (z_1, \ldots, z_L)$;
- (*uniqueness*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, the tuple $\mathbf{u}^O$ such that (3.25) holds is unique;
- (*closedness*) the map $(\mathbf{g}, \mathbf{u}^I) \mapsto \mathbf{u}^O$, with domain U and codomain $\mathcal{O}$, given as the unique solution of (3.25), defines a closed operator.

Definition 3.6.4 (**Semantics of PDAS**)**.**

(a) Let $K, L \in \mathbb{N}$ and $N = K + L$. Let $\mathcal{P}$ be a PDAS with K inputs and L outputs. Let $U \subseteq \mathcal{C} \times \mathcal{I}$ such that $\mathcal{P}$ is quasi-well-posed on U. We define the *solution* of $\mathcal{P}$ as the (partial) function

$$F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O};$$
$$F(\mathbf{g}, \mathbf{u}^I) = \mathbf{u}^O,$$

whose domain is U and where $\mathbf{u}^O$ is given by (3.25). We also say that $\mathcal{P}$ *generates* F on $U = \mathrm{dom}(F)$.

(b) A function $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O}$ is *PDAS-generable* if there exists a PDAS $\mathcal{P}$ such that $\mathcal{P}$ is quasi-well-posed on the domain of F and F is the solution of $\mathcal{P}$.

Remark 3.6.5. Observe that a PDAS describes a system of L equations on the variables $y_1, \ldots, y_K$, $z_1, \ldots, z_L$, which can be seen as an L-dimensional 'surface' on the N-dimensional space of variables. The notion of quasi-well-posedness allows us to define a semantics map $(y_1, \ldots, y_K) \mapsto (z_1, \ldots, z_L)$. This should be compared to the Implicit Function Theorem, which allows us to make such statements on Euclidean spaces (however, in our case the variables are not real but elements of $C(\mathbb{T}, \mathcal{X})$ instead). Hence, in principle, (local) well-posedness would be equivalent to some invertibility property on the *Jacobian* corresponding to the PDAS,

$$J(y_1, \ldots, y_K, z_1, \ldots, z_L) = \begin{bmatrix} \frac{\partial P_1}{\partial y_1} & \cdots & \frac{\partial P_1}{\partial y_K} & \frac{\partial P_1}{\partial z_1} & \cdots & \frac{\partial P_1}{\partial z_L} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial P_L}{\partial y_1} & \cdots & \frac{\partial P_L}{\partial y_K} & \frac{\partial P_L}{\partial z_1} & \cdots & \frac{\partial P_L}{\partial z_L} \end{bmatrix}.$$

We also remark that in a PDAS, there is a correspondence between the outputs and a subset of the initial conditions (namely, those for which $\beta = 0$), and so we have the following basic property (cf. Propositions 3.4.12 and 3.5.5).

Proposition 3.6.6. *If $F : \mathcal{C} \times \mathcal{I} \rightharpoonup \mathcal{O}$ is PDAS-generable, then* $\dim(\mathcal{O}) \leq \dim(\mathcal{C})$.

Our next two results will complete the cycle in Figure 3.10, showing that generable functions in each mode can be seen as projections of generable functions in the other modes.

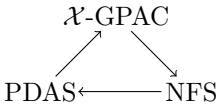


Figure 3.10: Cycle of main results.

Lemma 3.6.7 (**NFS-generable implies PDAS-generable**). *Any NFS-generable function is PDAS-generable.*

Proof. Any normal form equation is a partial differential algebraic equation where the variables occur with time (and space) derivatives of order at most 1; thus the initial conditions in an NFS are exactly those appearing as initial conditions in the corresponding PDAS; in other words, every NFS is a PDAS. □

Lemma 3.6.8 (**PDAS-generable implies $\mathcal{X}$-GPAC-generable**). *Let $F_P : \mathcal{C}_P \times \mathcal{I}_P \rightharpoonup \mathcal{O}_P$ be PDAS-generable with domain $U_P \subseteq \mathcal{C}_P \times \mathcal{I}_P$. Then there exists $F_G : \mathcal{C}_G \times \mathcal{I}_G \rightharpoonup \mathcal{M}_G \times \mathcal{O}_G$ which is $\mathcal{X}$-GPAC-generable with domain $U_G \subseteq \mathcal{C}_G \times \mathcal{I}_G$ with the following properties:*

- $\dim(\mathcal{I}_G) = \dim(\mathcal{I}_P) + 1$ *and* $\dim(\mathcal{C}_G) \geq \dim(\mathcal{C}_P)$*;*
- *for every $(\mathbf{g}, \mathbf{u}^I) \in \mathcal{C}_P \times \mathcal{I}_P$ such that $(\mathbf{g}, \mathbf{u}^I) \notin U_P$, we have $(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I, y) \notin U_G$ for any $\mathcal{X}$-vector $\mathbf{g}^*$ and any $\mathcal{X}$-stream y;*
- *for every $(\mathbf{g}, \mathbf{u}^I) \in \mathcal{C}_P \times \mathcal{I}_P$ such that $(\mathbf{g}, \mathbf{u}^I) \in U_P$, there exists a unique $\mathcal{X}$-vector $\mathbf{g}^*$ and $\mathcal{X}$-stream y such that $(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I, y) \in U_G$; moreover, there exists a unique $\mathcal{X}$-stream vector $\mathbf{u}^*$ such that*
$$F_G(\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I, y) = (\mathbf{u}^*, F_P(\mathbf{g}, \mathbf{u}^I)).$$

In other words, F_P is a projection of F_G.

The equality and inequality in the first bullet will be explained below.

Proof. Let $F_P : \mathcal{C}_P \times \mathcal{I}_P \rightharpoonup \mathcal{O}_P$ be PDAS-generable with domain $U_P \subseteq \mathcal{C}_P \times \mathcal{I}_P$. Denote by $\mathcal{P}$ the PDAS that generates F_P.

The important idea of the proof is understanding how to write partial differential algebraic equations with $\mathcal{X}$-GPAC modules. We start with channels for all variables $y_1, \ldots, y_K, z_1, \ldots, z_L$. To obtain derivatives in time, we include modules and connections as in Figure 3.11.

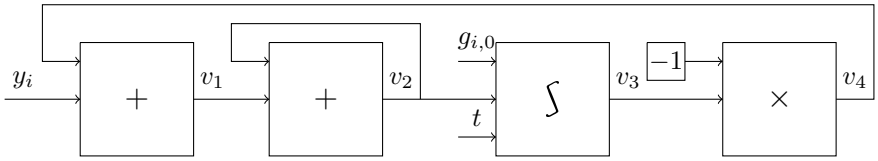


Figure 3.11: An $\mathcal{X}$-GPAC for computing time derivatives.

Note that the channels on Figure 3.11 must obey the system

$$\begin{cases} v_1 = y_i + v_4; \\ v_2 = v_1 + v_2; \\ v_3 = g_{i,0} + \int v_2 dt; \\ v_4 = -v_3; \\ y_i(0) = g_{i,0}; \end{cases} \Rightarrow \begin{cases} v_1 = 0; \\ v_2 = y_i'; \\ v_3 = y_i; \\ v_4 = -y_i; \end{cases}$$

so that we obtain y_i' in the channel labeled v_2. Observe that we must include an initial setting (in the above case, the initial value $y_i(0)$) every time we need to take a time derivative. We also need to include a channel carrying linear time $\mathbf{t} \in C^1([0, T], \mathcal{X})$. By composing several of these

modules, we can get all time derivatives that appear in $\mathcal{P}$, that is, we obtain $y_k^{(\beta)}$ for $k = 1, \ldots, K$, $\beta = 0, \ldots, \beta_k - 1$; and $z_\ell^{(\beta)}$ for $\ell = 1, \ldots, L$, $\beta = 0, \ldots, \beta_{K+\ell} - 1$.

Next, by successive applications of the differential module (see Figure 3.12) we obtain all the partial derivatives appearing in $\mathcal{P}$, which are of the form $\partial_x^\alpha y_k^{(\beta)}$ for $k = 1, \ldots, K$, $\alpha = 0, \ldots, \alpha_k$, $\beta = 0, \ldots, \beta_k - 1$; and $\partial_x^\alpha z_\ell^{(\beta)}$ for $\ell = 1, \ldots, L$, $\alpha = 0, \ldots, \alpha_{K+\ell}$, $\beta = 0, \ldots, \beta_{K+\ell} - 1$.

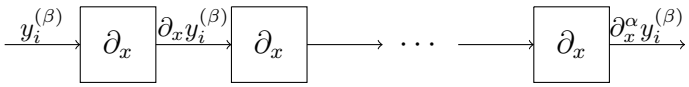


Figure 3.12: An $\mathcal{X}$-GPAC for computing spatial derivatives.

Next, we compute the polynomial expressions P_ℓ from the partial derivative terms using constants, multipliers and adders. Any term appearing in the polynomials P_ℓ is of the form $at_1^{d_1} \cdots t_n^{d_n}$, where $a \in \mathbb{R}$ (obtainable from a constant module), each t_i is either t (obtainable using the channel $\mathbf{t}$ carrying linear time) or some $\partial_x^\alpha y_k^{(\beta)}$, $\partial_x^\alpha z_\ell^{(\beta)}$, and thus can be obtained by a constant module and a sequence of multipliers. Each polynomial P_ℓ is a finite sum of such terms, and thus can be obtained by a sequence of adders.

Finally, to enforce the relation $P_\ell(t, y_1, \ldots, y_K, z_1, \ldots, z_L \ldots, \partial_x^{\alpha_1} y_1^{(\beta_1)}, \ldots, \partial_x^{\alpha_N} z_L^{(\beta_N)}) = 0$, we add z_ℓ to both sides of the equation and include an adder and feedback loop, as shown on Figure 3.13. We can then loop the channel labeled z_ℓ back to the start, enforcing it to be a mixed channel.

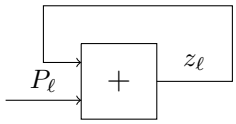


Figure 3.13: Feedback loop implementing $P_\ell = 0$.

Figure 3.14 illustrates the several steps of our construction, which results in an $\mathcal{X}$-GPAC $\mathcal{G}$. We must next address the underlying spaces of $\mathcal{G}$. The constant space is constructed with the initial settings of integrators, which only appear in the phase where we build time derivatives. For every original variable y_k which appears in $\mathcal{P}$ with time derivatives of order up to β_k, we have included the β_k initial settings $g_{k,\beta} = y_k^{(\beta)}(0)$, $0 \leq \beta < \beta_k$. Hence, the constant space of $\mathcal{G}$, denoted by $\mathcal{C}_G$, is an extension of $\mathcal{C}_P$, which only takes into account the initial settings associated with the output variables z_ℓ; therefore $\dim(\mathcal{C}_G) \geq \dim(\mathcal{C}_P)$.

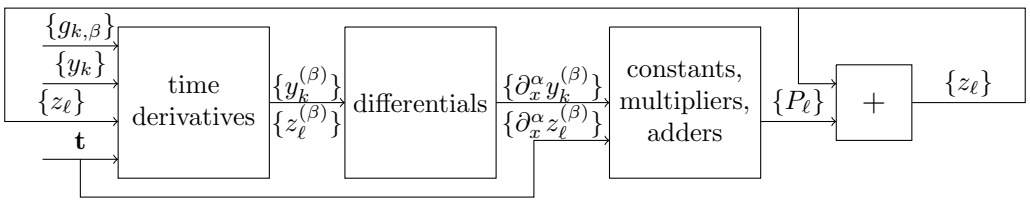


Figure 3.14: Construction of an $\mathcal{X}$-GPAC from a PDAS.

In regards to the input space, the original input variables y_k, $k = 1, \ldots, K$ appear in $\mathcal{G}$ as proper input channels. The only other proper input channel is linear time $\mathbf{t}$, which can be seen as an

'explicit time' inserted into the system. In this way, $\dim(\mathcal{I}_G) = \dim(\mathcal{I}_P) + 1$, which proves the first bullet.

Observe that the original output variables z_ℓ, $\ell = 1, \dots, L$, all appear in $\mathcal{G}$ as mixed channels because of the feedback loop that ensures $P_\ell = 0$. There are (potentially many) other mixed and proper output channels in $\mathcal{G}$, which carry the value of either partial derivatives $\partial_x^\alpha y_k^{(\beta)}$, $\partial_x^\alpha z_\ell^{(\beta)}$, multiplying terms $at_i^{d_i} \cdots t_n^{d_n}$ or sums of multiplying terms; hence $\dim(\mathcal{M}_G \times \mathcal{O}_G) \geq \dim(\mathcal{O}_P)$.

The next step is to find a suitable U_G for the domain of F_G. By quasi-well-posedness of $\mathcal{P}$, we know that, for every $(\mathbf{g}, \mathbf{u}^I) \in U_P$, there exists a unique $\mathbf{u}^O \in \mathcal{O}_P$ that solves $\mathcal{P}$. Thus, if we define $(\mathbf{g}^*, \mathbf{g})$ as the vector of initial conditions of all time derivatives of the input and output variables, which are of the form $y_k^{(\beta)}(0)$ for $k = 1, \dots, K$, $\beta = 0, \dots, \beta_k - 1$ and $z_\ell^{(\beta)}(0)$ for $\ell = 1, \dots, L$, $\beta = 0, \dots, \beta_{K+\ell} - 1$, we can infer that $\mathbf{g}^*$ depends uniquely on $(\mathbf{u}^I, \mathbf{u}^O)$, and thus it depends uniquely on $(\mathbf{g}, \mathbf{u}^I)$. With this in mind we define

$$U_G = \{(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) : (\mathbf{g}, \mathbf{u}^I) \in U_P \text{ and } (\mathbf{g}^*, \mathbf{g}) = (y_1, \dots, y_K^{(\beta_K - 1)}, z_1, \dots, z_L^{(\beta_{K+L}-1)}) \restriction_{t=0},$$
$$\text{where } (y_1, \dots, y_K) = \mathbf{u}^I \text{ and } (z_1, \dots, z_\ell) = \mathbf{u}^O = F_P(\mathbf{g}, \mathbf{u}^I)\}.$$

To define F_G, we need to define the value of all the mixed and output channels appearing in $\mathcal{G}$. As described above, these are either the output variables $z_1, \dots, z_L$, or uniquely obtained from the tuple $(\mathbf{u}^I, \mathbf{u}^O, \mathbf{t}) = (y_1, \dots, y_K, z_1, \dots, z_L, \mathbf{t})$ via partial derivatives, products and sums. If we let $\mathbf{u}^*$ denote the value of all the mixed and output channels which are not the output variables, then $\mathbf{u}^*$ depends uniquely on $(\mathbf{u}^I, \mathbf{u}^O, \mathbf{t})$ and thus also on $(\mathbf{g}, \mathbf{u}^I)$, so that we can define

$$F_G(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) = \begin{cases} (\mathbf{u}^*, F_P(\mathbf{g}, \mathbf{u}^I)) & \text{if } (\mathbf{g}, \mathbf{g}^*, \mathbf{u}^I, \mathbf{t}) \in U_G; \\ \text{undefined} & \text{otherwise.} \end{cases}$$

By this construction, U_G and F_G satisfy the second and third bullet points, and all is left is to show that F_G is the specification of $\mathcal{G}$ on U_G. By quasi-well-posedness of $\mathcal{P}$ on U_P and the above discussion, it is clear that for $(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) \in U_G$, the tuple $(\mathbf{u}^*, \mathbf{u}^O) \in \mathcal{O}_G$ that solves the equational specification of $\mathcal{G}$ exists and is unique; that is to say, $\mathbf{u}^O$ is given by $F_P(\mathbf{g}, \mathbf{u}^I)$ and $\mathbf{u}^*$ is obtained from $(\mathbf{u}^I, \mathbf{u}^O, \mathbf{t})$ via partial derivatives, produts and sums.

To prove closedness of F_G, consider a sequence[5] $(\mathbf{g}_n^*, \mathbf{g}_n, \mathbf{u}_n^I, \mathbf{t}) \in U_G$ such that

$$(\mathbf{g}_n^*, \mathbf{g}_n, \mathbf{u}_n^I, \mathbf{t}) \to (\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) \text{ and } F_G(\mathbf{g}_n^*, \mathbf{g}_n, \mathbf{u}_n^I, \mathbf{t}) \to (\mathbf{u}^*, \mathbf{u}^O);$$

by defining $(\mathbf{u}_n^*, \mathbf{u}_n^O) = F_G(\mathbf{g}_n^*, \mathbf{g}_n, \mathbf{u}_n^I, \mathbf{t})$, we have

$$(\mathbf{g}_n, \mathbf{u}_n^I) \to (\mathbf{g}, \mathbf{u}^I) \text{ and } F_P(\mathbf{g}_n, \mathbf{u}_n^I) = \mathbf{u}_n^O \to \mathbf{u}^O.$$

By closedness of F_P, we have

$$(\mathbf{g}, \mathbf{u}^I) \in U_P \text{ and } F_P(\mathbf{g}, \mathbf{u}^I) = \mathbf{u}^O.$$

Now $(\mathbf{g}_n^*, \mathbf{g}_n)$ corresponds to the values of the variables in $(\mathbf{u}_n^I, \mathbf{u}_n^O)$ and some of their derivatives at $t = 0$. By closedness of the differential operator, we can take limits and conclude that $(\mathbf{g}^*, \mathbf{g})$ corresponds to the values of the variables in $(\mathbf{u}^I, \mathbf{u}^O)$ and their derivatives at $t = 0$; thus, $(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) \in U_G$. Also, since partial derivatives, products and sums are closed operators, we infer that $(\mathbf{u}_n^*, \mathbf{u}_n^O) \to F_G(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t})$, so that $F_G(\mathbf{g}^*, \mathbf{g}, \mathbf{u}^I, \mathbf{t}) = (\mathbf{u}^*, \mathbf{u}^O)$, which concludes the proof. □

[5] Since the last component of any tuple in U_G must be linear time $\mathbf{t}$, we only need to consider sequences for the other three subtuples.

Example 3.6.9. We provide an example of the construction in the previous Lemma by applying it to the one-dimensional heat equation, which can be written as the following PDAS,

$$u' = \partial_x^2 u, \quad u(0) = g.$$

To produce a solution to the heat equation, we consider, for example, a square-integrable initial condition $g \in C(\mathbb{R}) \cap L^2(\mathbb{R})$, for which the solution can be obtained via the heat kernel, that is, $F_P : \mathcal{X} \rightharpoonup C^1([0,T],\mathcal{X})$ is given by $F_P(g) = u$, where

$$u(t,x) = \int_{\mathbb{R}} K(t, x-y) g(y) dy, \quad K(t,x) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}.$$

Let us construct an $\mathcal{X}$-GPAC with specification F_G such that F_P is a projection of F_G. We start with a channel for the variable u. Using the constructions on Figures 3.11 and 3.12 we obtain channels with the derivatives u', $\partial_x u$ and $\partial_x^2 u$. We can then construct the partial differential algebraic expression $P(u, u', \partial_x u, \partial_x^2 u) = u' - \partial_x^2 u$ using one constant module, one multiplier module and one adder module. Finally, we include an adder and feedback loop as in Figure 3.13 and loop the variable u to the beginning. The final $\mathcal{X}$-GPAC can be seen in Figure 3.15.

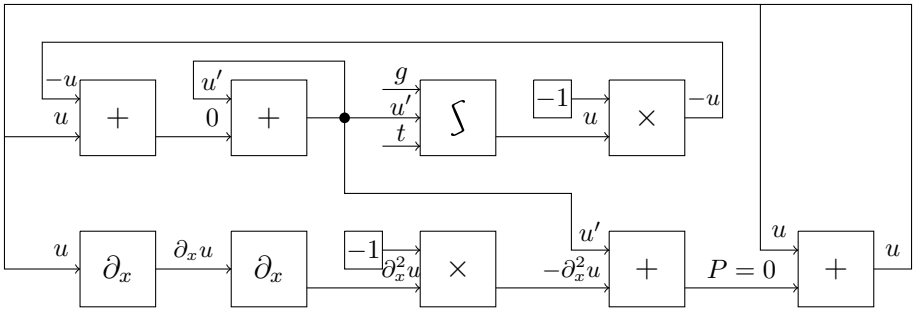


Figure 3.15: Construction of an $\mathcal{X}$-GPAC from the heat equation.

We can see that the $\mathcal{X}$-GPAC has one additional input $\mathbf{t}$, which carries linear time. It should also be noted that the heat equation has only one variable u, whereas the $\mathcal{X}$-GPAC has a total of twelve channels and thus twelve variables (of course, most of those are redundant). Therefore, any specification of the $\mathcal{X}$-GPAC must be of type $\mathcal{X}^1 \times C^1([0,T],\mathcal{X})^1 \rightharpoonup C^1([0,T],\mathcal{X})^{11}$.

We can also see that if $F_P(g) = u$ produces a solution to the heat equation, then there is a unique tuple $\mathbf{u}$ of values that satisfy $\Phi(g, \mathbf{t}, \mathbf{u}) = \mathbf{u}$, where Φ is the induced operator of the $\mathcal{X}$-GPAC in Figure 3.15, g is the initial condition of the integrator and $\mathbf{t}$ is the input channel (linear time). In other words, we can construct a specification $F_G(g, \mathbf{t}) = \mathbf{u}$ of the $\mathcal{X}$-GPAC such that F_P is a projection of F_G, as expected.

It should be clear from this example that the construction in Lemma 3.6.8 is universal (in the sense that it works for any PDAS) but is not necessarily optimal (in the sense that it does not add a minimal number of new variables). In fact, one could construct a much simpler $\mathcal{X}$-GPAC (with only three modules!) that generates solutions to the heat equation, as in Figure 3.16. The reader can verify that F_P is also a projection of the specification of this $\mathcal{X}$-GPAC.

Finally we can state and prove our main result.

Theorem 17 ($\mathcal{X}$**-GPAC Characterization Theorem**). *Let*

$$F : \mathcal{X}^j \times C^1([0,T],\mathcal{X})^k \rightharpoonup C^1([0,T],\mathcal{X})^\ell.$$

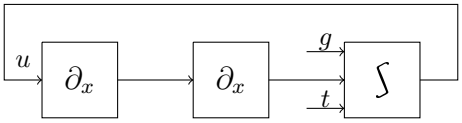


Figure 3.16: An $\mathcal{X}$-GPAC that generates solutions to the heat equation.

The following are equivalent:

- *F is the projection of an $\mathcal{X}$-GPAC-generable function;*
- *F is the projection of an NFS-generable function;*
- *F is the projection of a PDAS-generable function;*

Proof. We prove each implication:

($\mathcal{X}$-GPAC $\Rightarrow$ NFS) Let F be the projection of an $\mathcal{X}$-GPAC-generable function F_G. By Lemma 3.5.7, F_G is the projection of an NFS-generable function F_N, and so F is a projection of F_N.

(NFS $\Rightarrow$ PDAS) Similar to the previous case, but use Lemma 3.6.7 instead.

(PDAS $\Rightarrow$ $\mathcal{X}$-GPAC) Similar to the previous case, but use Lemma 3.6.8 instead. □

3.7 The Multityped GPAC

The rest of this chapter is devoted to a more abstract study of the $\mathcal{X}$-GPAC. The addition of $C^1([0,T],\mathcal{X})$ as a new space of data channels suggests a generalization of the GPAC model to manipulate data in many-sorted data types. As we mentioned in Section 3.3, we could consider other function spaces, such as

$$\mathcal{X} = C^p(\Omega) \quad \text{or} \quad \mathcal{X} = H^p(\Omega),$$

where Ω is a domain in $\mathbb{R}^n$. Another possible direction would be to consider modules operating on multiple data types, which could be written as

$$\Phi : \tau_1^{\ell_1} \times \ldots \times \tau_n^{\ell_n} \to \tau;$$

in other words, we can define a model of computation where different channels may carry different types of data.

This direction of research would have a strongly technical aspect. As a starting point, we shall extend the known model by taking $\mathcal{X}$ to be any complete metric vector space and assuming the existence of four different types of channels (cf. beginning of Section 3.1):

- $\mathbb{R}$-*scalar* channels, which carry a constant $k \in \mathbb{R}$;
- $\mathcal{X}$-*scalar* channels, which carry a constant $x \in \mathcal{X}$;
- $\mathbb{R}$-*stream* channels, which carry a stream $a \in C^1([0,T],\mathbb{R})$;
- $\mathcal{X}$-*stream* channels, which carry a stream $u \in C^1([0,T],\mathcal{X})$.

We make the important observation that adding *scalar* channels that carry constants will give us some extra freedom and enable some operations that technically were not present in Shannon's original construction. We must also redefine our basic modules.

Definition 3.7.1 (**Basic $\mathcal{X}$-modules**). The *basic $\mathcal{X}$-modules* are defined as follows:

- for any $k \in \mathbb{R}$, the *k-constant* module has zero inputs and one $\mathbb{R}$-scalar output. It outputs the constant k. Similarly, for any $x \in \mathcal{X}$, the *x-constant* module has zero inputs and one $\mathcal{X}$-scalar output. It outputs the constant x;
- the *$\mathbb{R}$-adder* module has two $\mathbb{R}$-stream inputs and one $\mathbb{R}$-stream output. For inputs a and b, it outputs the sum $a + b$. Similarly, the *$\mathcal{X}$-adder* module has two $\mathcal{X}$-stream inputs and one $\mathcal{X}$-stream output. For inputs u and v, it outputs the sum $u + v$;
- the *$\mathbb{R}$-scalar-$\mathbb{R}$-multiplier* module has one $\mathbb{R}$-scalar input, one $\mathbb{R}$-stream input and one $\mathbb{R}$-stream output. For inputs k and a, it outputs the product ka. Similarly, the *$\mathbb{R}$-scalar-$\mathcal{X}$-multiplier* module has one $\mathbb{R}$-scalar input, one $\mathcal{X}$-stream input and one $\mathcal{X}$-stream output. For inputs k and u, it outputs the product ku;
- the *$\mathbb{R}$-integrator* module has one $\mathbb{R}$-scalar input, two $\mathbb{R}$-stream inputs and one $\mathbb{R}$-stream output. For inputs k, a and b, it outputs the Lebesgue-Stieltjes integral $k + \int a db$. Similarly, there are two *$\mathcal{X}$-integrator* modules, each having one $\mathcal{X}$-scalar input, one $\mathbb{R}$-stream input, one $\mathcal{X}$-stream input and one $\mathcal{X}$-stream output. For inputs x, a and u, one outputs the Lebesgue-Stieltjes integral $x + \int a du$ and the other outputs the Lebesgue-Stieltjes integral $x + \int u da$;

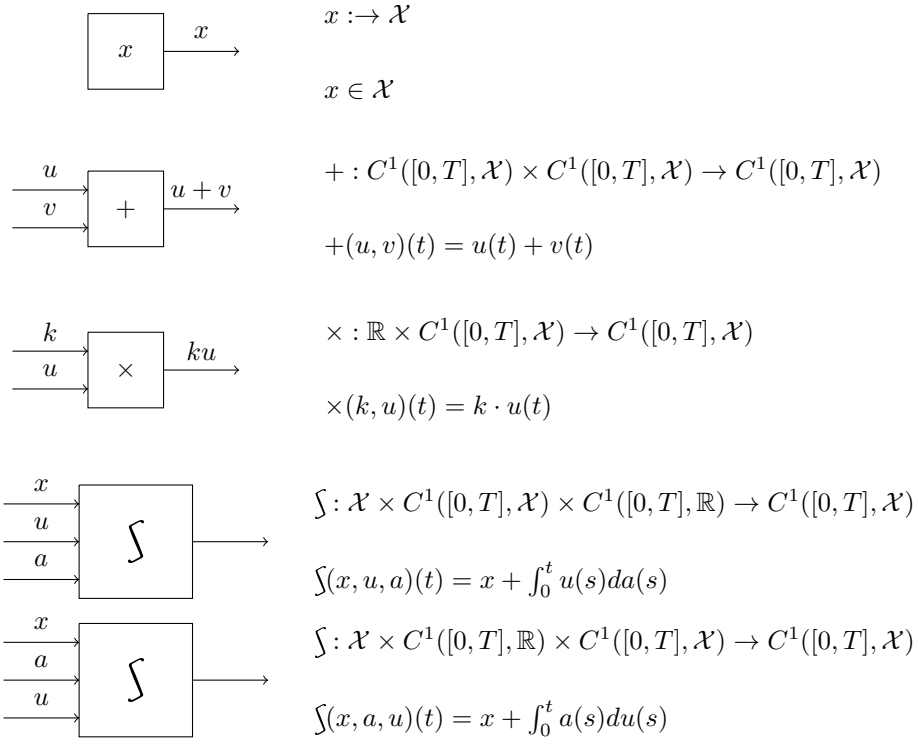


Figure 3.17: Some basic modules in an $\mathcal{X}$-GPAC.

Remark 3.7.2. Notice that we did not define an integrator module with two $\mathcal{X}$-stream inputs. The reason is that, at this level of generality, there may not be a product operation of the type

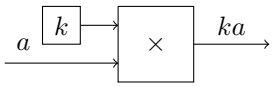


Figure 3.18: Derivation of Shannon scalar multiplier modules.

$\mathcal{X} \times \mathcal{X} \to \mathcal{X}$. Notice also that the scalar multiplier and integrators in Definition 3.7.1 have an extra scalar input channel compared to their Shannon counterparts in Definition 3.1.1. This can be considered as a generalization, since we can represent the old modules via the new ones, as stated in the following proposition.

Proposition 3.7.3. *The Shannon basic modules (Definition 3.1.1) can be derived from the $\mathcal{X}$-basic modules (Definition 3.7.1).*

Proof. The Shannon adder is given by the $\mathbb{R}$-adder. Figure 3.18 shows how to derive the Shannon multiplier from a k-constant and an $\mathbb{R}$-scalar-$\mathbb{R}$-multiplier. The Shannon integrator is given by the $\mathbb{R}$-integrator, making sure the $\mathbb{R}$-scalar input is a proper input. The Shannon constant can be derived from a 0-constant, a 1-constant, an $\mathbb{R}$-scalar-$\mathbb{R}$-multiplier and an $\mathbb{R}$-integrator; we leave the diagram to the proof of Proposition 3.8.6 after we introduce constant streamers.

□

3.8 Module derivation and channel contraction

In this section we present a few more additional modules which we find useful to understand the power of $\mathcal{X}$-GPACs. Moreover, we present two operations on GPACs to increase or reduce the number of modules.

The main motivation is that some operations which one may consider 'fundamental' are not captured in the basic modules from Figures 3.1 and 3.17. For example, the multiplication of two $\mathbb{R}$-scalars,

$$\times(k, \ell) = k\ell,$$

or the multiplication of two $\mathbb{R}$-streams,

$$\times(a, b)(t) = a(t)b(t),$$

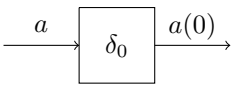
are not directly obtained in any of the previous modules. In regards to the multiplication of $\mathbb{R}$-streams, both Shannon and Pour-El argue that this can be obtained by the other basic modules, due to the relation

$$a(t)b(t) = a(0)b(0) + \int_0^t a(s)db(s) + \int_0^t b(s)da(s), \tag{3.26}$$

that holds for any $a, b \in C^1([0, T], \mathbb{R})$. However, the first term, $a(0)b(0)$, cannot be obtained generally from streams a and b via the basic modules, unless in specific circumstances (for example, if we restrict a and b to the space of streams such that $a(0) = 0$). Therefore, there is no GPAC that computes the right hand side of (3.26), if only the basic modules are considered.

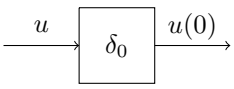
To solve this issue, one could simply introduce a new module to compute the multiplication of two $\mathbb{R}$-streams, and include it in our definition of GPAC (indeed, this is what we did in Definition 3.4.4). This would however raise other questions, as there are many varieties of modules which we would then need to include, one by one, as will be clear from the rest of this section. The path we have chosen is instead to include one additional type of module, namely *evaluation modules*, from which all these operations can be derived.

Definition 3.8.1 (**Initial evaluator modules**)**.** The $\mathbb{R}$-*initial evaluator* has one $\mathbb{R}$-stream input and one $\mathbb{R}$-scalar output. For an input a, it outputs the initial value $a(0)$. Similarly, the $\mathcal{X}$-*initial evaluator* has one $\mathcal{X}$-stream input and one $\mathcal{X}$-scalar output. For an input u, it outputs the initial value $u(0)$.



$$\delta_0 : C^1([0,T],\mathbb{R}) \to \mathbb{R}$$

$$\delta_0(a) = a(0)$$

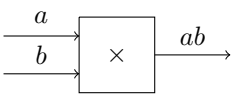


$$\delta_0 : C^1([0,T],\mathcal{X}) \to \mathcal{X}$$

$$\delta_0(u) = u(0)$$

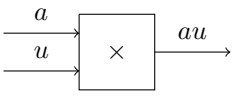
Figure 3.19: Initial evaluator modules.

Definition 3.8.2 (**Stream multiplier modules**)**.** The $\mathbb{R}$-*stream multiplier* modules have one $\mathbb{R}$-stream input, one $\mathbb{R}$-stream or $\mathcal{X}$-stream input and one $\mathbb{R}$-stream or $\mathcal{X}$-stream output, respectively. For inputs a, and b or u respectively, they output the products ab or au respectively.



$$\times : C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$\times(a,b)(t) = a(t)b(t)$$



$$\times : C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathcal{X}) \to C^1([0,T],\mathcal{X})$$

$$\times(a,u)(t) = a(t)u(t)$$

Figure 3.20: $\mathbb{R}$-stream multiplier modules.

Proposition 3.8.3. *The $\mathbb{R}$-stream multiplier modules can be derived from the initial evaluator and basic modules.*

Proof. Figure 3.21 shows how to derive the $\mathbb{R}$-stream-$\mathbb{R}$-multiplier (that is, with two $\mathbb{R}$-stream inputs and one $\mathbb{R}$-stream output) using two $\mathbb{R}$-initial evaluators, an $\mathbb{R}$-scalar-$\mathbb{R}$-multiplier, a 0-constant, two $\mathbb{R}$-integrators and an $\mathbb{R}$-adder. A similar proof works for the case $C^1([0,T],\mathbb{R}) \times C^1([0,T],\mathcal{X}) \to C^1([0,T],\mathcal{X})$ as well.

□

Example 3.8.4. We show the usefulness of the $\mathbb{R}$-stream multiplier by using it to build an *inverter*, a partially defined function given by

$$F : \mathbb{R} \times C^1([0,T],\mathbb{R}) \rightharpoonup C^1([0,T],\mathbb{R}); \quad F(k,b)(t) = \frac{k}{1 + k(b(t) - b(0))}. \tag{3.27}$$

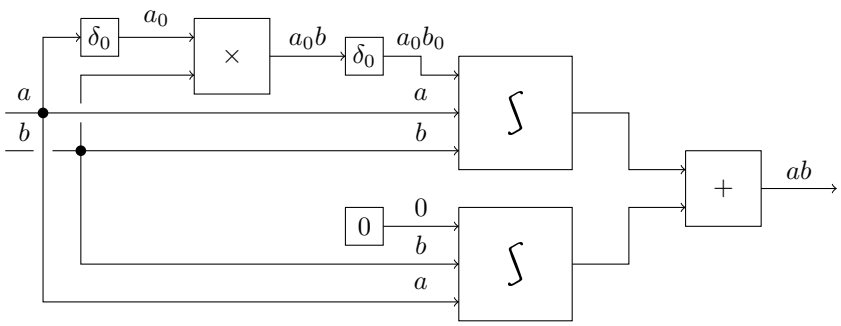


Figure 3.21: Derivation of an $\mathbb{R}$-stream multiplier module.

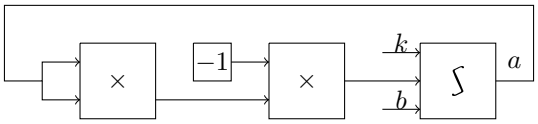


Figure 3.22: A GPAC generating the inverter functional.

Let us show that F is GPAC-generable by considering the GPAC in Figure 3.22.

This GPAC induces a system of four equations on six variables, which is reducible to a single equation on the channels labeled k, a and b, given by

$$a'(t) = -a(t)^2 b'(t), \quad a(0) = k;$$

after some calculations we find the unique solution to be

$$a(t) = \frac{k}{1 + k(b(t) - b(0))}.$$

Therefore, F is a (component of a) GPAC-generable partial function; its domain is given by

$$D(F) = \{(k, b) \in \mathbb{R} \times C(\mathbb{T}, \mathbb{R}) : k = 0 \text{ or } b(t) \neq b(0) - 1/k \text{ for all } t \in \mathbb{T}\}.$$

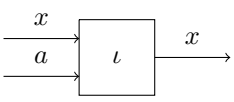
It is worth noticing that, when $k = 1$ and $b = \mathbf{t}$ is linear time, the corresponding solution is $a(t) = \frac{1}{1+t}$, which provides an example for a GPAC-generable rational function.

The next module we consider can be thought of as the inverse of the evaluator.

Definition 3.8.5 (**Streamer modules**)**.** The $\mathbb{R}$-*constant streamer* module has one $\mathbb{R}$-scalar input, one $\mathbb{R}$-stream input and one $\mathbb{R}$-stream output. For inputs k, a, it outputs the constant stream k. Similarly, the $\mathcal{X}$-*constant streamer* module has one $\mathcal{X}$-scalar input, one $\mathcal{X}$-stream input and one $\mathcal{X}$-stream output. For inputs x, u, it outputs the constant stream x.

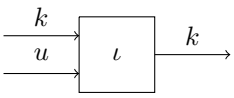
Proposition 3.8.6. *The $\mathbb{R}$-constant streamer module can be derived from the basic modules. The $\mathcal{X}$-constant streamer module can be derived from the basic modules using an additional $\mathbb{R}$-stream input.*

Proof. Figure 3.24 shows how to derive constant streamers using a 0-constant, an $\mathbb{R}$-scalar multiplier



$$\iota : \mathbb{R} \times C^1([0,T],\mathbb{R}) \to C^1([0,T],\mathbb{R})$$

$$\iota(k,a)(t) = k$$



$$\iota : \mathcal{X} \times C^1([0,T],\mathcal{X}) \to C^1([0,T],\mathcal{X})$$

$$\iota(x,u)(t) = x$$

Figure 3.23: Constant streamer modules.

and an $\mathbb{R}$-integrator. The main idea of the proof is the trivial relation

$$k = k + \int_0^t 0 da(s).$$

Observe also that, for the $\mathcal{X}$-constant streamer, an additional $\mathbb{R}$-stream input is used; that is, the $\mathcal{X}$-constant streamer only has two inputs but we required a third input for the GPAC in the right-hand side of Figure 3.24. In the current framework this extra channel is necessary, since the $\mathcal{X}$-integrator module has an $\mathbb{R}$-stream input. However, if a product operation of type $\mathcal{X} \times \mathcal{X} \to \mathcal{X}$ exists (cf. Remark 3.7.2), then we can use the corresponding $\mathcal{X}$-integrator module (and thus we would not require such an additional channel).

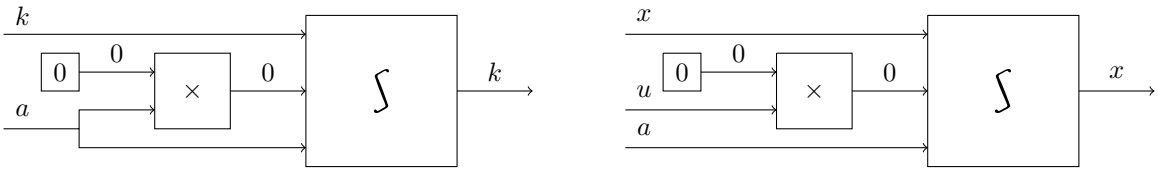


Figure 3.24: Derivation of constant streamer modules.

□

We also observe that the Shannon constant module can be obtained from a constant streamer, letting k be the output of a 1-constant (cf. proof of Proposition 3.7.3).

Clearly, once we have constant streamers and initial evaluators, we can derive addition and multiplication of scalars from their stream counterparts.

Definition 3.8.7 (**Scalar operation modules**)**.** The *scalar adder* and *scalar multiplier* modules are defined as in Definition 3.7.1, but replacing every stream channel by its corresponding scalar channel.

Proposition 3.8.8. *The scalar adders and scalar multipliers can be derived from the constant streamers, initial evaluators, and basic modules, using an additional stream input.*

Proof. Figure 3.26 shows how to derive the scalar adder and scalar multiplier of type $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$. The same idea works for the remaining cases.

□

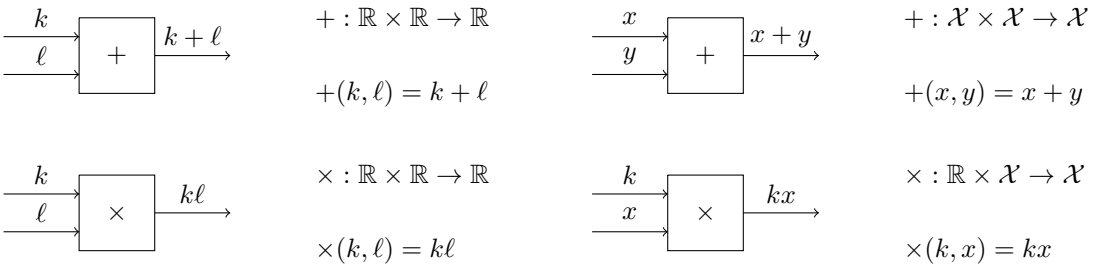


Figure 3.25: Scalar adder and scalar multiplier modules.

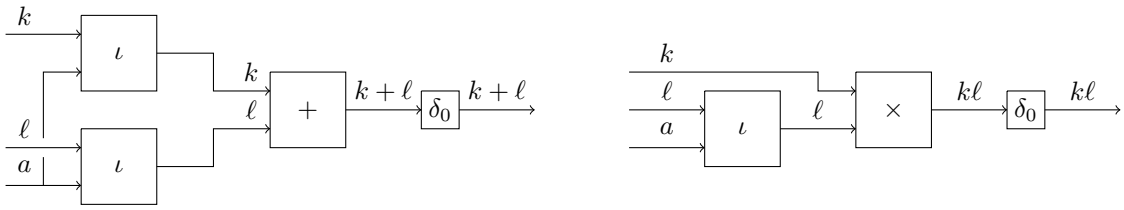


Figure 3.26: Derivation of scalar adder and scalar multiplier modules.

All the previous propositions describe how a variety of modules can be derived from the basic modules and the initial evaluators. In an effort to formalize this notion, we introduce the concept of derived module.

Definition 3.8.9 (**Derived module**)**.** A *derived operation* is a functional $F : \mathcal{I} \to \mathcal{O}$ that can be obtained as a composition of the basic modules and the initial evaluators. In other words, there exists a GPAC $\mathcal{G}$ built only with basic modules and initial evaluators, with induced operator $\Phi_0 : \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}$ such that

- $\mathcal{G}$ is an acyclic graph;
- Φ_0 has the same input and output spaces as F;
- $\mathcal{G}$ is well-posed on the whole input space, and F is the projection of the specification F_0 of $\mathcal{G}$ onto its output space,
$$F = \pi_{\mathcal{O}} \circ F_0.$$

A *derived module* is a module that specifies a derived operation.

Remark 3.8.10. To clarify the definition of derived operation, let $\mathcal{G}$ be an acyclic GPAC. Since all the basic modules define total functions and $\mathcal{G}$ has no loops, it is clear that $\mathcal{G}$ is well-posed on the whole input space $\mathcal{I}$ (that is, all mixed and output channels have well-defined values for any valuation of the input channels). We can then define the induced operator Φ_0 (which can be 'read off' $\mathcal{G}$) and the specification F_0 (which solves the fixed point equation). Finally, the derived operation F is obtained from F_0 by 'ignoring' the mixed channels and considering only the output channels. Formally, we have

$$\Phi_0 : \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}, \quad F_0 : \mathcal{I} \to \mathcal{M} \times \mathcal{O}, \quad F : \mathcal{I} \to \mathcal{O};$$

and for all $\mathbf{in} \in \mathcal{I}$, $\mathbf{out} \in \mathcal{O}$, we have

$$\begin{aligned} \Phi_0(\mathbf{in}, \mathbf{mix}) = (\mathbf{mix}, \mathbf{out}), \text{ for some } \mathbf{mix} \in \mathcal{M} \text{ iff} \\ F_0(\mathbf{in}) = (\mathbf{mix}, \mathbf{out}), \text{ for some } \mathbf{mix} \in \mathcal{M} \text{ iff} \\ F(\mathbf{in}) = \mathbf{out}. \end{aligned}$$

Example 3.8.11. The stream multipliers, constant streamers, scalar adders and scalar multipliers are derived modules.

Definition 3.8.12 (**GPAC reducibility**)**.** Let $\mathcal{G}$ and $\mathcal{G}'$ be GPACs, not necessarily exclusively composed of basic modules. Let $\mathcal{I}$, $\mathcal{I}'$, $\mathcal{M}$, $\mathcal{M}'$, $\mathcal{O}$, $\mathcal{O}'$ be their corresponding input, mixed and output spaces. We say that $\mathcal{G}$ is *reducible* to $\mathcal{G}'$ if

- $\mathcal{G}$ and $\mathcal{G}'$ have the same input and output spaces, $\mathcal{I} = \mathcal{I}'$, $\mathcal{O} = \mathcal{O}'$, and the mixed space of $\mathcal{G}'$ is a subspace of the mixed space of $\mathcal{G}$, $\mathcal{M}' \subseteq \mathcal{M}$;
- for any open set $U \subseteq \mathcal{I}$, $\mathcal{G}$ is well-posed on U if and only if $\mathcal{G}'$ is well-posed on U;
- for any open set $U \subseteq \mathcal{I}$ on which $\mathcal{G}$, $\mathcal{G}'$ are well-posed, if $F : \mathcal{I} \rightharpoonup \mathcal{M} \times \mathcal{O}$ is the specification of $\mathcal{G}$ and $F' : \mathcal{I} \rightharpoonup \mathcal{M}' \times \mathcal{O}$ is the specification of $\mathcal{G}'$, then F' is the projection of F onto the subspace $\mathcal{M}' \times \mathcal{O}$,
$$F' = \pi_{\mathcal{M}' \times \mathcal{O}} \circ F.$$

Theorem 18 (**Expansion of derived modules**)**.** *For any GPAC $\mathcal{G}$ composed of derived modules, there is a GPAC $\mathcal{G}'$ composed of basic modules and initial evaluators such that $\mathcal{G}'$ is reducible to $\mathcal{G}$.*

Proof. Given a GPAC $\mathcal{G}$ composed of derived modules, replace each derived module by the corresponding composition of basic modules and initial evaluators (inserting additional mixed channels as necessary), until a GPAC $\mathcal{G}'$ composed of basic modules and initial evaluators is obtained. It is then clear that $\mathcal{G}'$ is reducible to $\mathcal{G}$. □

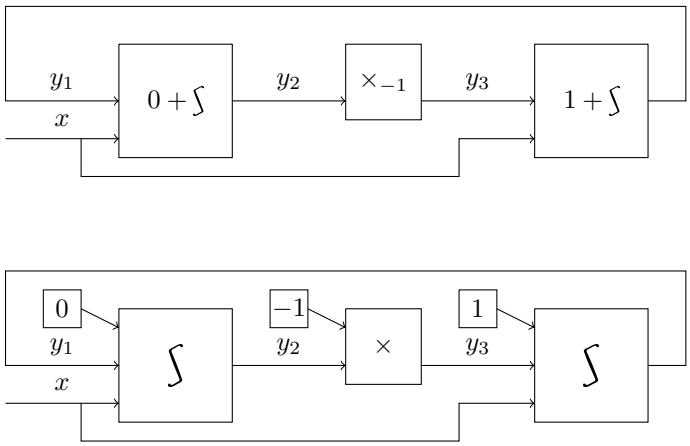
Example 3.8.13. Figure 3.27 shows an example of a GPAC composed of three derived modules that computes sine and cosine functions, and the corresponding GPAC composed of basic modules as in Theorem 18.

Theorem 18 describes a process in which we transform a GPAC with derived modules into a GPAC with basic modules and initial evaluators, by increasing the number of modules and channels. We now present the reverse process, by which we can contract channels to obtain a GPAC with fewer modules and channels.

Definition 3.8.14 (**Contractible channel**)**.** A mixed channel of a GPAC is said to be *contractible* if it does not connect an input with an output of the same module, in other words, if it is the output of a module M and the input of one or more modules, neither of which is M. A GPAC is said to be *contraction-free* if it does not have contractible channels.

Theorem 19 (**Contraction of GPACs**)**.** *For any GPAC $\mathcal{G}$, there is a contraction-free GPAC $\mathcal{G}'$ such that $\mathcal{G}$ is reducible to $\mathcal{G}'$.*

Proof. Let $\mathcal{G}$ be a GPAC with a contractible channel y. Assume that y is the output of some module M, where M has inputs $x_1, \ldots, x_k$ and generates an operation $y = \Phi(x_1, \ldots, x_k)$. Assume that y connects as an input to modules $M_1, \ldots, M_n$. Consider the GPAC $\mathcal{G}'$ obtained from $\mathcal{G}$ in the following way (see Figure 3.28):



$$y_1(t) = 1 + \int_0^t y_3(s)x'(s)ds;$$

$$y_2(t) = \int_0^t y_1(s)x'(s)ds;$$

$$y_3(t) = -y_2(t);$$

Figure 3.27: Two GPACs that specify trigonometric functions; the first one is composed of three derived modules; the second one is composed of six basic modules, being reducible to the first one; the solution of the system is $y_1(t) = \cos(x(t) - x(0))$, $y_2(t) = \sin(x(t) - x(0))$, $y_3(t) = -\sin(x(t) - x(0))$.

- remove module M and replace modules $M_1, \ldots, M_n$ with modules $M_1', \ldots, M_n'$;
- connect all the inputs $x_1, \ldots, x_k$ of module M as inputs to modules $M_1', \ldots, M_n'$;
- if M_i specifies an operation $z_i = \Phi_i(y, y_1, \ldots, y_m)$, then let M_i' specify an operation $z_i = \Phi_i'(x_1, \ldots, x_k, y_1, \ldots, y_m)$ via the composition

$$\Phi_i'(x_1, \ldots, x_k, y_1, \ldots, y_m) = \Phi_i(\Phi(x_1, \ldots, x_k), y_1, \ldots, y_m).$$

In this way, we obtain a GPAC $\mathcal{G}'$ such that $\mathcal{G}$ is reducible to $\mathcal{G}'$ and $\mathcal{G}'$ has one less module and at least one less contractible channel than $\mathcal{G}$. Iterating this procedure a finite number of times, we eventually arrive at a contraction-free GPAC.

□

Remark 3.8.15. For any GPAC $\mathcal{G}$, the procedure indicated in the proof of Theorem 19 is ensured to terminate and produce a contraction-free GPAC $\mathcal{G}'$. However, $\mathcal{G}'$ may not be the unique contraction-free GPAC to which $\mathcal{G}$ is reducible; in fact, different choices of contractible channels may yield different contraction-free GPAC, as the following example demonstrates.

Example 3.8.16. Returning to Example 3.8.13 and Figure 3.27, we see that there are three contractible channels, labeled y_1, y_2 and y_3. By contracting any two of these three channels, we arrive at a contraction-free GPAC. Therefore, there are three possible reductions of the original GPAC to a contraction-free GPAC, as shown in Figure 3.29.

3.9 Contracting GPACs and contracting operators

In the original ideas of Tucker and Zucker concerning analog networks [TZ07], the notions of causal and contracting operators are presented and studied in great detail. Their main case study is the mass-spring-damper system, which models a mass M suspended by a spring with stiffness K and damping coefficient D and subject to a force f (which is a function of time). The goal of this

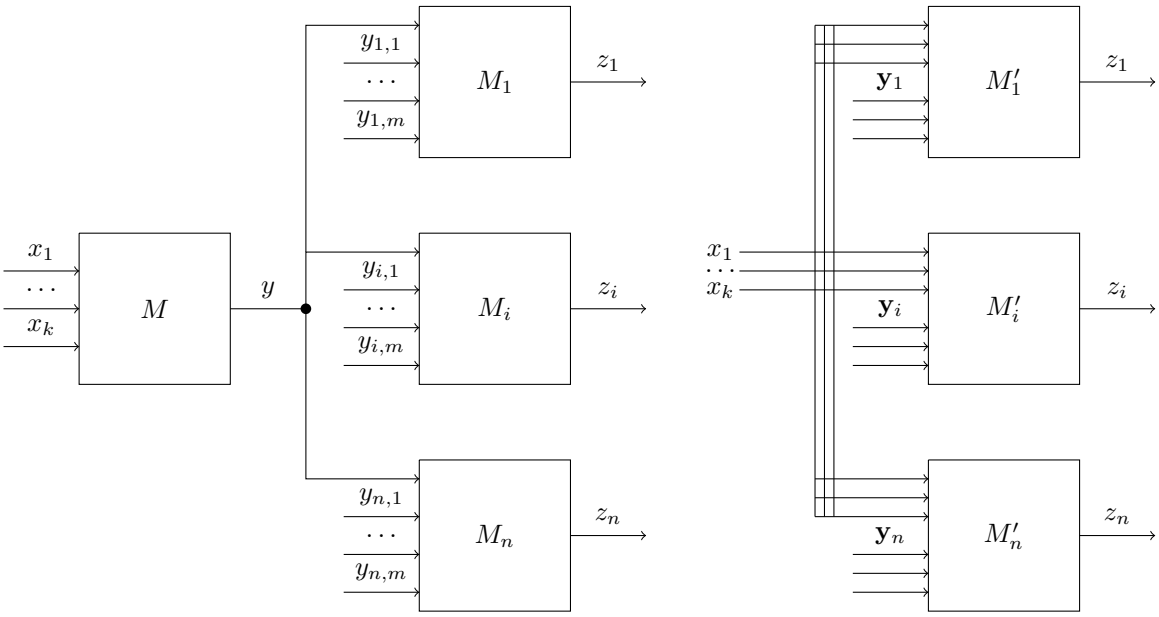


Figure 3.28: Schematic representation of channel contraction.

section is to briefly illustrate how the notion of contracting operator and the spring-mass-damper system can be expressed in our framework.

Remark 3.9.1. We remark that, on a multityped GPAC, we can define metrics on each of the associated spaces $\mathcal{I}$, $\mathcal{M}$ and $\mathcal{O}$. These are induced by the 'standard' metrics on $\mathbb{R}$, $\mathcal{X}$, $C^1([0,T],\mathbb{R})$ and $C^1([0,T],\mathcal{X})$. Moreover, for Lemma 3.9.4 we shall assume in addition that the metric is induced from a p-norm with $1 \leq p < \infty$; in other words, if $(\mathcal{X}_1, d_1), \ldots (\mathcal{X}_n, d_n)$ are metric spaces, we define a metric on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ by[6]

$$d((x_1,\ldots,x_n),(y_1,\ldots,y_n)) = (d(x_1,y_1)^p + \ldots + d(x_n,y_n)^p)^{1/p}. \tag{3.28}$$

Definition 3.9.2 (**Contracting operator**)**.** Let $\mathcal{G}$ be a GPAC with induced operator $\Phi : \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}$.

- For $\mathbf{y}^I \in \mathcal{I}$, Φ is said to be *contracting* at $\mathbf{y}^I$ if $\pi_{\mathcal{M}} \circ \Phi(\mathbf{y}^I, \cdot)$ is a contraction mapping with respect to the metric on $\mathcal{M}$ (cf. Definition 2.1.4). In other words, there exists $\lambda \in [0,1)$ such that, for any $\mathbf{y}_1^M, \mathbf{y}_2^M \in \mathcal{M}$, writing

  $$\Phi(\mathbf{y}^I, \mathbf{y}_1^M) = (\tilde{\mathbf{y}}_1^M, \mathbf{y}_1^O);$$

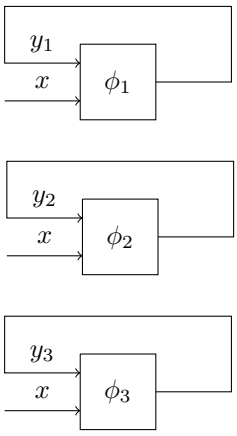
  $$\Phi(\mathbf{y}^I, \mathbf{y}_2^M) = (\tilde{\mathbf{y}}_2^M, \mathbf{y}_2^O);$$

  then

  $$d_{\mathcal{M}}(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M) \leq \lambda d_{\mathcal{M}}(\mathbf{y}_1^M, \mathbf{y}_2^M).$$

- For an open set $U \subseteq \mathcal{I}$, Φ is said to be *(uniformly) contracting* on U if $\pi_{\mathcal{M}} \circ \Phi(\mathbf{y}^I, \cdot)$ is a contraction mapping for each $\mathbf{y}^I \in U$; and moreover the modulus of contractivity λ does not

[6] We can also consider $p = \infty$, in which case we obtain the maximum norm.



$$\phi_1(y_1, x)(t) = 1 + \int_0^t \left(- \int_0^s y_1(\tau) x'(\tau) d\tau \right) x'(s) ds;$$

$$\phi_2(y_2, x)(t) = \int_0^t \left(1 + \int_0^s - y_2(\tau) x'(\tau) d\tau \right) x'(s) ds;$$

$$\phi_3(y_3, x)(t) = - \int_0^t \left(1 + \int_0^s y_3(\tau) x'(\tau) d\tau \right) x'(s) ds;$$

Figure 3.29: Three possible contraction-free reductions of the GPAC in Figure 3.27; the solution to each system is $y_1 = \cos(x(t))$, $y_2 = -\sin(x(t))$ and $y_3 = \sin(x(t))$.

depend on $\mathbf{y}^I \in U$.

Proposition 3.9.3. *Let $\mathcal{G}$ be a GPAC with induced operator $\Phi : \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O}$.*

1. *Let $\mathbf{y}^I \in \mathcal{I}$ and suppose that Φ is contracting at $\mathbf{y}^I$. Then there exist unique $\mathbf{y}^M \in \mathcal{M}$, $\mathbf{y}^O \in \mathcal{O}$ such that*
$$\Phi(\mathbf{y}^I, \mathbf{y}^M) = (\mathbf{y}^M, \mathbf{y}^O). \tag{3.29}$$
2. *Let $U \subseteq \mathcal{I}$ be open and suppose that Φ is uniformly contracting on U. Then $\mathcal{G}$ is well-posed on U.*

Proof. Both results follow easily from the Banach fixed point theorem (Theorem 2) and its proof. Note that, in claim 2, we need to prove that the functional $\mathbf{y}^I \mapsto (\mathbf{y}^M, \mathbf{y}^O)$ giving solutions to (3.29) is continuous; since the modulus of contraction λ is uniform on U, this is achieved by expressing such functional as the limit of an iteration scheme with uniform modulus of convergence. □

Lemma 3.9.4 (**Channel contraction preserves contractivity**). *Let $\mathcal{G}$ and $\mathcal{G}'$ be GPACs with induced operators Φ, Φ' and suppose that $\mathcal{G}'$ is obtained from $\mathcal{G}$ by contraction on one of its contractible channels (as per the proof of Theorem 19). Assume also that the metrics on the underlying spaces are as in Remark 3.9.1. Then, if Φ is contracting, so is Φ'.*

Proof. The key ingredient in this proof is understanding how to express Φ' (the induced operator of the contracted GPAC) in terms of Φ (the induced operator of the original GPAC) and the module function whose output is the channel being contracted. Let us begin by writing the induced operators of $\mathcal{G}$ and $\mathcal{G}'$ as of type

$$\Phi : \mathcal{I} \times \mathcal{M} \to \mathcal{M} \times \mathcal{O};$$
$$\Phi' : \mathcal{I} \times \mathcal{M}' \to \mathcal{M}' \times \mathcal{O};$$

moreover, let y denote the contractible channel and $\mathcal{M}_C$ denote its underlying space, so that we can write $\mathcal{M} = \mathcal{M}_C \times \mathcal{M}'$.

Now consider $\mathbf{y}^I \in \mathcal{I}$ such that Φ is contracting at $\mathbf{y}^I$, say with modulus of contractiveness λ. To show that Φ' is also contracting at $\mathbf{y}^I$, let us take $\mathbf{y}_1^M, \mathbf{y}_2^M, \tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M, \mathbf{y}_1^O, \mathbf{y}_2^O$ with

$$\Phi'(\mathbf{y}^I, \mathbf{y}_1^M) = (\tilde{\mathbf{y}}_1^M, \mathbf{y}_1^O);$$

$$\Phi'(\mathbf{y}^I, \mathbf{y}_2^M) = (\tilde{\mathbf{y}}_2^M, \mathbf{y}_2^O).$$

Let Φ_C be the component of Φ associated to the contractible channel y. In other words, Φ_C is the function associated to the module in $\mathcal{G}$ whose output is y (cf. Figure 3.28). Since y is a contractible channel, Φ_C does not depend on the variable y; in other words, we can think of Φ_C as a function of type $\mathcal{I} \times \mathcal{M}' \to \mathcal{M}_C$; moreover, Φ' can be expressed in terms of Φ and Φ_C as $(\Phi_C, \Phi') = \Phi(\cdot, \Phi_C, \cdot)$. To be precise, we mean the following: if $y_1 = \Phi_C(\mathbf{y}^I, \mathbf{y}_1^M)$ and $y_2 = \Phi_C(\mathbf{y}^I, \mathbf{y}_2^M)$, then

$$\Phi(\mathbf{y}^I, y_1, \mathbf{y}_1^M) = (y_1, \Phi'(\mathbf{y}^I, \mathbf{y}_1^M)) = (y_1, \tilde{\mathbf{y}}_1^M, \mathbf{y}_1^O);$$

$$\Phi(\mathbf{y}^I, y_2, \mathbf{y}_2^M) = (y_2, \Phi'(\mathbf{y}^I, \mathbf{y}_2^M)) = (y_2, \tilde{\mathbf{y}}_2^M, \mathbf{y}_2^O).$$

We now use the fact that Φ is contracting at $\mathbf{y}^I$ to conclude that

$$d_{\mathcal{M}}((y_1, \tilde{\mathbf{y}}_1^M), (y_2, \tilde{\mathbf{y}}_2^M)) \leq \lambda d_{\mathcal{M}}((y_1, \mathbf{y}_1^M), (y_2, \mathbf{y}_2^M)). \tag{3.30}$$

Observe that y_1 and y_2 appear on both sides of the inequality; the next step is to make them disappear in order to obtain the desired bound on the remaining channels. Using the assumption that $d_{\mathcal{M}}$ is induced from a p-norm, we write

$$d_{\mathcal{M}}((y_1, \tilde{\mathbf{y}}_1^M), (y_2, \tilde{\mathbf{y}}_2^M))^p = d(y_1, y_2)^p + d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p;$$

$$d_{\mathcal{M}}((y_1, \mathbf{y}_1^M), (y_2, \mathbf{y}_2^M))^p = d(y_1, y_2)^p + d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p;$$

putting these into (3.30), we deduce as follows:

$$d(y_1, y_2)^p + d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p \leq \lambda^p (d(y_1, y_2)^p + d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p);$$

$$d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p \leq \lambda^p d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p + (\lambda^p - 1) d(y_1, y_2)^p;$$

$$d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p \leq \lambda^p d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M)^p;$$

$$d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M) \leq \lambda d(\tilde{\mathbf{y}}_1^M, \tilde{\mathbf{y}}_2^M),$$

where on the third step we use the fact that $(\lambda^p - 1) < 0$. We conclude that Φ' is contracting at $\mathbf{y}^I$.

Finally, if Φ is uniformly contracting on $U \subseteq \mathcal{I}$, then it is contracting at each $\mathbf{y}^I \in U$ with some modulus of contractivity independent of $\mathbf{y}^I$. The previous argument then shows that Φ' is also contracting at each $\mathbf{y}^I \in U$ with the same modulus of contractivity, so that Φ' is uniformly contracting on U as well. □

Example 3.9.5. The rest of this section is dedicated to the study of the spring-mass-damper system, as presented in [TZ07, Jam12]. This system can be expressed as the ODE

$$Mx''(t) + Dx'(t) + Kx(t) = f(t), \tag{3.31}$$

where M, D, K are positive constants. The system also has initial conditions $x(0) = x_0$ and $x'(0) = v_0$ on the *displacement* and *velocity*.

By introducing new variables v, a for the *velocity* and *acceleration*, we can write the second-order equation as a first-order system, which can be integrated to obtain

$$a(t) = \frac{1}{M} f(t) - \frac{D}{M} v(t) - \frac{K}{M} x(t);$$

$$v(t) = \int_0^t a(s)ds + v_0;$$
$$x(t) = \int_0^t v(s)ds + x_0.$$

In this way, the mass-spring-damper system can be represented by the analog network of Figure 3.30.

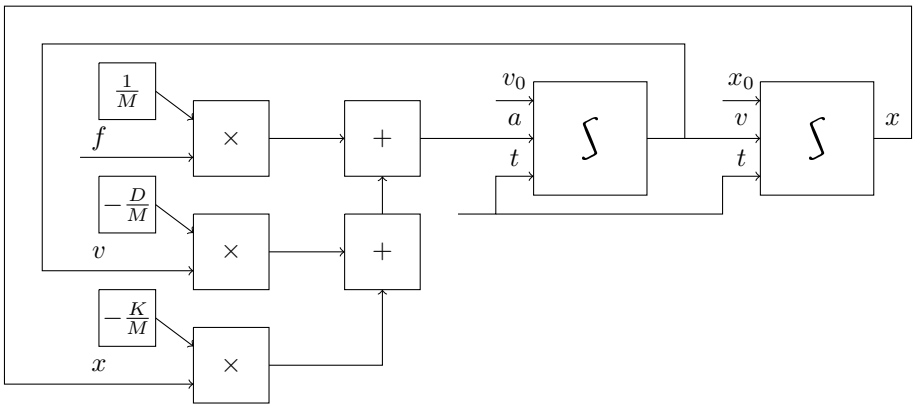


Figure 3.30: A GPAC comprised of basic modules for the mass-spring-damper system.

The GPAC presented in Figure 3.30 has a too large number of ten modules. As noted in [TZ07], the network can be simplified by combining the constant, adder and scalar multiplier modules into a single module that performs a weighted sum. In our framework, this corresponds to contracting the GPAC on most of its channels. The resulting network (with only three modules) is presented in Figure 3.31. We remark that its induced operator is given by

$$\Phi_{TZ} : \mathbb{R}^2 \times C^1([0,T],\mathbb{R})^5 \to C^1([0,T],\mathbb{R})^3$$
$$\Phi_{TZ}(x_0, v_0, f, x, v, a, \mathbf{t}) = \left(\frac{1}{M}f - \frac{D}{M}v - \frac{K}{M}x, \int ads + v_0, \int vds + x_0\right). \tag{3.32}$$

Tucker and Zucker study the contractiveness of the operator defined in (3.32) and their results can be expressed in our framework as follows.

Proposition 3.9.6. *Suppose that $M > \max\{K, 2D\}$. Then there exists a sufficiently small $T \in \mathbb{R}+$ such that, for all $x_0, v_0 \in \mathbb{R}, f \in C^1([0,T],\mathbb{R})$, the operator Φ_{TZ} defined in* (3.32) *is contracting at* $(x_0, v_0, f, \mathbf{t})$.

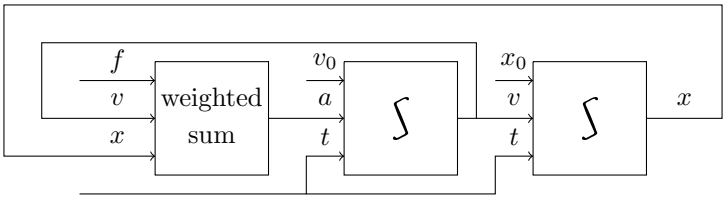


Figure 3.31: Simplified network for the mass-spring-damper system.

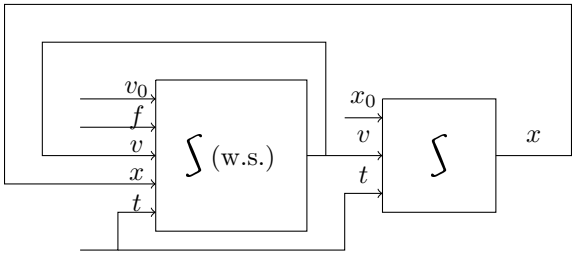


Figure 3.32: Further simplified network for the mass-spring-damper system.

Proof. See [TZ07]. □

As evidenced from the above proposition, and as shown by James in [Jam12], the GPAC in Figure 3.31 is not the most robust formulation for the mass-spring-damper system. By this we mean that there are choices of parameters that violate the condition $M > \max\{K, 2D\}$ and for which the operator Φ_{TZ} defined in (3.32) is not contracting. However, it is not the case that the spring-mass-damper system behaves erratically in some choices of parameter space. It turns out that we can consider a slight modification of the GPAC in Figure 3.31 whose induced operator is contracting for all choices of M, K, D, as our intuition would suggest. This revision was first proposed in [Jam12]; in our framework, it can be expressed as the contraction of the GPAC in Figure 3.31 on the acceleration channel a. The resulting network (with only two modules) is presented in Figure 3.32. Its induced operator is given by

$$\Phi_J : \mathbb{R}^2 \times C^1([0,T],\mathbb{R})^4 \to C^1([0,T],\mathbb{R})^2$$
$$\Phi_J(x_0, v_0, f, x, v, \mathbf{t}) = \left(\int \frac{1}{M} f - \frac{D}{M} v - \frac{K}{M} x ds + v_0, \int v ds + x_0 \right). \tag{3.33}$$

Proposition 3.9.7. *There exists a sufficiently small $T \in \mathbb{R}+$ such that, for all $M, K, D \in \mathbb{R}^+$, $x_0, v_0 \in \mathbb{R}$, $f \in C^1([0,T],\mathbb{R})$, the operator Φ_J defined in* (3.33) *is contracting at $(x_0, v_0, f, \mathbf{t})$.*

Proof. See [Jam12, Section 3.3]. □

In summary, we see that, given a GPAC comprised of basic modules (such as the one in Figure 3.30), we can simplify it by a sequence of channel contractions, obtaining other GPACs related by the notion of reducibility (such as the ones in Figures 3.31 and 3.32). These reducibilities preserve contractivity, that is, if the original GPAC has a contracting induced operator, then so do its reductions (Lemma 3.9.4). On the other hand, if the original GPAC does not have a contracting induced operator, we may still hope that, after performing a sufficient number of channel contractions, we may arrive at a contracting induced operator, as illustrated by the spring-mass-damper system.

3.10 Discussion

In this chapter we presented the $\mathcal{X}$-GPAC as a generalization of the Shannon GPAC. We considered the case $\mathcal{X} = C(\mathbb{R})$ and, in the first part of the chapter, we introduced a differential module that computes spatial derivatives. Theorem 17 is evidence that our model of computation provides a suitable generalization of the original work on the GPAC. We can thus think of solutions to certain differential equations as outputs or fixed points of certain analog networks. We have also seen that (quasi)well-posedness conditions play an important role in this study.

A possible direction for research would be to consider function spaces other than $C(\mathbb{R})$, such as

$$\mathcal{X} = C^p(\Omega) \quad \text{or} \quad \mathcal{X} = H^p(\Omega),$$

where Ω is, for example, a domain in $\mathbb{R}^n$, either unbounded (such as $\Omega = \mathbb{R}^n$) or bounded (such as $\Omega = [0,1]^n$), and H^p denotes Sobolev spaces. In the case when Ω is bounded, we may further restrict our space $\mathcal{X}$ to functions with prescribed behaviour on the boundary, such as Dirichlet boundary conditions ($f = 0$ on $\partial\Omega$). These examples were presented in Section 3.3. We believe that such a direction could allow us to make interesting connections with the field of partial differential equations, where such spaces are ubiquitous.

While on the topic of partial differential equations, we repeat here the disclaimer presented in the introduction. We hope that our results show how the theory of PDEs can be applied into the theory of computable analysis, and in particular the GPAC model. The reverse direction, i.e. using the GPAC to 'solve more difficult or more efficiently' PDEs, is outside the scope of our study. As a philosophical remark, our equivalence result can be interpreted as follows.

> The task of finding whether a general GPAC is well-posed (i.e. 'computes some function') is *equally difficult* as finding whether a general (algebraic) PDE is well-posed (i.e. 'has some solution').

We also started a direction of research in considering modules that operate on multiple data types, which could be written as

$$\Phi : \tau_1^{\ell_1} \times \ldots \times \tau_n^{\ell_n} \to \tau;$$

in this way we obtained channels of different types and defined a notion of *multityped GPAC*, and corresponding *many-typed analog networks*. There is a strong technical aspect in following this direction, and we have only studied some basic concepts (such as module derivation, channel contraction and reducibility). However, we feel that this avenue seems promising and it could lead to a model of computation on many-sorted algebras as studied by Tucker and Zucker [TZ00, TZ07, TZ14].

There is also an important difference between the formalism adopted by other authors in the study of the GPAC and ours. Usually, the GPAC has been used to study the computability of functions of type $\mathbb{R} \to \mathbb{R}$; for example, a list of generable functions appears in [BGP16] that includes exponentials, logarithms, inverses, sine, cosine and arctangent. However, technically speaking, the model of GPAC studied in this thesis considers the computability of higher-order functionals, for example of type[7] $(\mathbb{R} \to \mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$; see, for example, Definition 3.4.11. The usual functions of interest, having type $\mathbb{R} \to \mathbb{R}$, can be obtained in our framework as the output of a GPAC-generable function for linear time input $\mathbf{t}$ and / or suitable real inputs.

A disadvantage of this formalism is that the usual functions of interest (that is, the functions of type $\mathbb{R} \to \mathbb{R}$ prevalent in other models of analog computation such as computable analysis or type-2 theory of effectivity) do not arise so 'cleanly'; one must introduce, say, a linear time input and express them as *outputs* of GPAC-generable functionals. However, there are two important advantages in considering higher-type functionals. First, we gain some expressivity by being able to study richer types of generable functions (in other words, computation over $(\mathbb{R} \to \mathbb{R}) \to (\mathbb{R} \to \mathbb{R})$ is objectively more expressive than computation over $\mathbb{R} \to \mathbb{R}$); second, we are able to express generable functionals as solutions of fixed point problems in an intuitive manner, which may allow us to apply fixed point techniques to study computability in continuous spaces (in some sense, this was the goal of Chapter 2).

[7] To be precise, we should write $C([0,T],\mathbb{R})^n \to C([0,T],\mathbb{R})^m$, but we can consider $n = m = 1$ for simplicity and take $C([0,T],\mathbb{R})$ as a subspace of $\mathbb{R} \to \mathbb{R}$.

Chapter 4

The limit GPAC and approximability

In this chapter we present a second extension to the Shannon GPAC model, presented in the previous chapter. In particular, we wish to incorporate the procedure of taking limits into our model of analog networks. In abstract terms, one may want to define a class of 'computable' elements $\mathcal{C}$ such that

$$\text{If } f \in \mathcal{C}, \text{ then } \lim f \in \mathcal{C}.$$

Of course, part of the problem is understanding what kinds of 'limit' we are allowed to consider. Usually, in computability theory on continuous spaces, we must demand that limits be 'effective', in the sense that the modulus of convergence is known *a priori* and thus we can effectively obtain an approximation to the limit within a prescribed precision. The notion of limit must also agree with the topology of the underlying space, which can be induced by a metric, a norm, or a family of pseudonorms. Thus if $\mathcal{X}$ is a function space we may be interested in 'uniform' or 'locally uniform' as opposed to 'pointwise' limits.

We begin by introducing *discrete* channel types, that is, channels which either assume discrete values or that are defined at discrete points in time. We then define the notions of Cauchy sequence, Cauchy stream and effective convergence. With those ingredients, we are able to consider a new module that takes (discrete or continuous) limits and yet another extension to the Shannon GPAC, which we call LGPAC. In the latter part of the chapter, we show how to generate some non-differentially algebraic functions, such as the gamma and Riemann zeta functions, which are our main motivation for including limits.

We briefly summarize the original content of this chapter. It is important to remark that the idea of introducing approximability into the GPAC model is not new and can attributed to Graça, [Gra04]. In particular, the paper [BCGH07] provides a notion of GPAC-computability also based on limits, remarkably showing an equivalence with the class of computable functions on a compact interval. However, we claim that (to the best of our knowledge) the approach of including a *module* that performs limits is original. In our framework, approximability is incorporated on the GPAC model itself, and not just in the way we define the GPAC semantics. This is how we suggest our results be contrasted to those of Bournez, Campagnolo, Graça, Hainry and other authors. Therefore, the main original content of this chapter consists in the introduction of limit modules (Definition 4.2.4) and the notion of LGPAC (Remark 4.2.5); the computability of the gamma and Riemann zeta functions (Theorems 20 and 21) can be seen as applications of this theory.

4.1 Discrete channel types

We recall the multityped GPAC whose definition was sketched in Section 3.7. We considered a complete metric vector space $\mathcal{X}$ and the following types of channels:

- $\mathbb{R}$-*scalar* channels, which carry a constant $k \in \mathbb{R}$;
- $\mathcal{X}$-*scalar* channels, which carry a constant $x \in \mathcal{X}$;
- $\mathbb{R}$-*stream* channels, which carry a stream $a \in C^1([0,T],\mathbb{R})$;
- $\mathcal{X}$-*stream* channels, which carry a stream $u \in C^1([0,T],\mathcal{X})$.

In this chapter we will briefly consider *discrete* channel types. In that sense, our model can be seen as a hybrid between digital and analog computation. The addition of more channels will undoubtedly increase the difficulty of studying the power of the GPAC; but we make the important remark that these discrete channel types are not essential to the main purpose of this chapter, which is to generate some non-differentially algebraic functions. Therefore these are included only as an illustration. In any case, here are the further channel types we may wish to consider:

- $\mathbb{N}$-*scalar* channels, which carry a constant $k \in \mathbb{N}$;
- $\mathbb{N}$-*sequence* channels, which carry a sequence $\{k_n\} \in \mathbb{N}^{\mathbb{N}}$;
- $\mathbb{R}$-*sequence* channels, which carry a sequence $\{k_n\} \in \mathbb{R}^{\mathbb{N}}$;
- $\mathcal{X}$-*sequence* channels, which carry a sequence $\{g_n\} \in \mathcal{X}^{\mathbb{N}}$.

We remark that the channel type corresponding to $\mathbb{N}$-streams is not necessary, since any continuous function of type $\mathbb{T} \to \mathbb{N}$ must be constant.

4.2 The limit operator and the limit GPAC

Let us make precise what we mean by effective limit. If we take $\mathcal{X}$ to be a complete metric space with a metric d, then

- a sequence $\{g_n\} \in \mathcal{X}^{\mathbb{N}}$ is a *Cauchy sequence* whenever

  for all $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for $m, n \in \mathbb{N}$ with $m, n \geq N$ one has $d(g_m, g_n) < \epsilon$;

- a stream $u \in C(\mathbb{T}, \mathcal{X})$ is a *Cauchy stream* whenever

  for all $\epsilon > 0$ there exists $T \in \mathbb{T}$ such that for $s, t \in \mathbb{T}$ with $s, t \geq T$ one has $d(u(s), u(t)) < \epsilon$;

To write the effective version of these limits, we begin by replacing the existential quantifiers with functions on the precision ϵ. A possible approach is given in the following definitions.

Definition 4.2.1 (Moduli of convergence).

1. A *discrete modulus of convergence* is a nondecreasing function $N : \mathbb{N} \to \mathbb{N}$.
2. A *continuous modulus of convergence* is a nondecreasing, nonnegative function $T \in C(\mathbb{T}, \mathbb{R})$.

Remark 4.2.2 (**Effective moduli of convergence**)**.** Both definitions of moduli of convergence can be effectivized in an intuitive manner. To effectivize the notion of discrete modulus of convergence $N : \mathbb{N} \to \mathbb{N}$, we can require that N be computable (in the traditional sense). To effectivize the notion of continuous modulus of convergence $T \in C(\mathbb{T}, \mathbb{R})$, we can require that T be GPAC-generable. In the latter case we may further specify what type of GPAC we are interested in: either the Shannon GPAC, the multityped GPAC from Section 3.7 (with or without the differential module) or the limit GPAC which we will develop in this chapter. However, for most of the time we will desire T to be a somewhat "simple" function, such as a monomial, an exponential, or a chain of exponentials, in which case the notion of Shannon GPAC-generability suffices. In fact, we expect the computational richness of the construction to be in the stream for which we are taking limits, but not on the modulus of convergence itself.

Definition 4.2.3 (**Effective limits on metric spaces**)**.**

1. Let N be a discrete modulus of convergence and $\{g_n\} \in \mathcal{X}^{\mathbb{N}}$. Then $\{g_n\}$ is an *N-convergent Cauchy sequence* if

$$\text{for all } \nu \in \mathbb{N}, \text{ for all } m, n \in \mathbb{N} \text{ with } m, n \geq N(\nu) \text{ one has } d(g_m, g_n) < 2^{-\nu}.$$

2. Let T be a continuous modulus of convergence and $u \in C(\mathbb{T}, \mathcal{X})$. Then u is a *T-convergent Cauchy stream* if

$$\text{for all } \tau \in \mathbb{T}, \text{ for all } s, t \in \mathbb{T} \text{ with } s, t \geq T(\tau) \text{ one has } d(u(s), u(t)) < 2^{-\tau}.$$

3. A sequence $\{g_n\} \in \mathcal{X}^{\mathbb{N}}$ is called an *effective Cauchy sequence* if there is an effective discrete modulus of convergence N such that $\{g_n\}$ is N-convergent.

4. A stream $u \in C(\mathbb{T}, \mathcal{X})$ is called an *effective Cauchy stream* if there is an effective continuous modulus of convergence T such that u is T-convergent.

An example of a modulus of convergence is given by the identity function, either discrete ($\mathrm{id} : \mathbb{N} \to \mathbb{N}$) or continuous ($\mathrm{id} \in C(\mathbb{T}, \mathbb{R})$). We note that any effective Cauchy sequence may be replaced by an id-convergent Cauchy sequence via a composition with its modulus of convergence; in other words, if $\{g_n\}$ is an N-convergent Cauchy sequence, then $\{g_{N(n)}\}$ is an id-convergent Cauchy sequence. Similarly, an effective Cauchy stream may be replaced by an id-convergent Cauchy stream. Thus we may assume, for convenience, that the modulus of convergence for a given effective limit is given by the identity map.

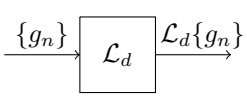
The next step is to introduce a *limit operator*, and again we may do this in a discrete or continuous manner.

Definition 4.2.4 (**Limit modules**)**.**

1. For the data type $\mathcal{X}$, there is a *discrete limit module* with one input of type $\mathcal{X}^{\mathbb{N}}$ and one output of type $\mathcal{X}$. For input $\{g_n\}$, it outputs the id-convergent limit $\lim_{n\to\infty} g_n$ (if it exists).

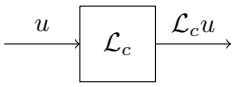
2. For the data type $\mathcal{X}$, there is a *continuous limit module* with one input of type $C(\mathbb{T}, \mathcal{X})$ and one output of type $\mathcal{X}$. For input u, it outputs the id-convergent limit $\lim_{t\to\infty} u(t)$ (if it exists).

A few comments are in order. Firstly, it should be clear that the limit modules define partial-valued operators; they are only defined for those sequences in $\mathcal{X}^{\mathbb{N}}$ (or those functions in $C(\mathbb{T}, \mathcal{X})$) that have an id-convergent limit. Secondly, the choice of the identity as the 'canonical' modulus of



$$\mathcal{L}_d : \mathcal{X}^{\mathbb{N}} \rightharpoonup \mathcal{X}$$

$$\mathcal{L}_d\{g_n\} = \lim_{n\to\infty} g_n$$

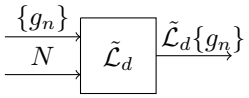


$$\mathcal{L}_c : C(\mathbb{T}, \mathcal{X}) \rightharpoonup \mathcal{X}$$

$$\mathcal{L}_c u = \lim_{t\to\infty} u(t)$$

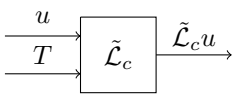
Figure 4.1: Limit modules.

convergence allows us to specify the limit operator as a one-input, one-output module. A different approach could be taken, in which a *two-input limit module* is considered, having one input for the sequence (or stream) and another input for the discrete (or continuous) modulus of convergence, such as in Figure 4.2.



$$\tilde{\mathcal{L}}_d : \mathcal{X}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}} \rightharpoonup \mathcal{X}$$

$$\tilde{\mathcal{L}}_d(\{g_n\}, N) = \lim_{n\to\infty} g_n$$



$$\tilde{\mathcal{L}}_c : C(\mathbb{T}, \mathcal{X}) \times C(\mathbb{T}, \mathbb{R}) \rightharpoonup \mathcal{X}$$

$$\tilde{\mathcal{L}}_c(u, T) = \lim_{t\to\infty} u(t)$$

Figure 4.2: Two-input limit modules.

Since, as explained above, any effective limit may be converted to an id-convergent limit via a composition with the modulus of convergence, we can derive the two-input limit module from the one-input limit module using a composition, as depicted on Figure 4.3.

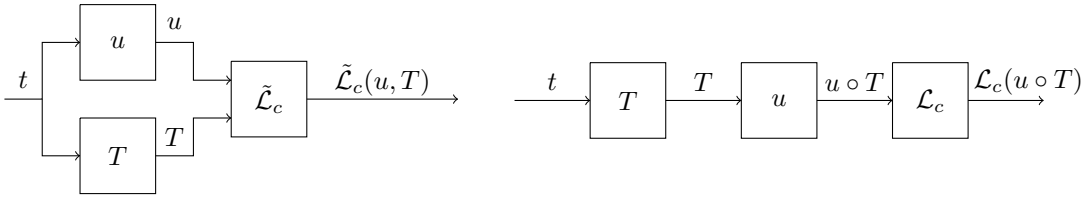


Figure 4.3: Derivation of the two-input continuous limit module; the discrete case is done similarly.

Remark 4.2.5 (**Limit GPAC**)**.** After abstracting the operation of taking limits as a module, we can sketch the definition of *limit GPAC* (or LGPAC). This can be achieved similarly as in Definitions 3.1.5 and 3.4.4. The notions of induced operator, well-posedness, LGPAC semantics and LGPAC-generability follow as well. Since the construction just mimics what was done in Chapter 3, we refrain from giving a precise definition of the LGPAC. Moreover, there are some non-obvious choices we would have to make to give that definition:

- Should we include the discrete channel types from Section 4.1 and the discrete limit module from Definition 4.2.4?
- Should we opt for the one-input or two-input limit modules?
- What notion of effective modulus of convergence should we consider (cf. Remark 4.2.2)?

Clearly, as we increase the variety (in both channel types and modules) of our construction, we get more inclusive models of computation, but finding characterization results becomes increasingly difficult and technical. Keeping in mind that our goal is to compute some non-differentially algebraic functions such as the gamma function and the Riemann zeta function, we can limit our construction to the minimum that makes that goal achievable. As it shall be seen in Sections 4.5 and 4.6, this can be accomplished by considering only continuous channel types and GPAC-generable continuous moduli of convergence.

4.3 Infinite speedup, infinite slowdown

The composition presented in Figure 4.3 can be thought of as a *time speedup* by T (or *slowdown*, if T grows slower than the identity). The goal of this section is to observe that infinite speedups can also be expressed in our model. Thus, the choice of limit $t \to \infty$ is not the only possibility, as one may consider limits of the form $t \to T^-$ for any positive time $T \in \mathbb{T}$. In order to see this, we consider the following functions that continuously map the interval $[0, 1)$ to $[0, \infty)$ and vice versa.

Proposition 4.3.1. *The following functions are Shannon GPAC-generable:*

- *(infinite speedup)* $t \mapsto \frac{t}{1-t}$, *with domain* $[0, 1)$ *and range* $[0, \infty)$*;*
- *(infinite slowdown)* $t \mapsto \frac{t}{1+t}$, *with domain* $[0, \infty)$ *and range* $[0, 1)$.

Proof. Recall the inverter functional $\Phi : (k, b) \mapsto a(t) = \frac{k}{1+k(b(t)-b(0))}$ constructed in Example 3.8.4. The function $s_\uparrow(t) = \frac{1}{1-t}$ can be obtained as the output of Φ with $k = 1$ and $b(t) = -t$. The function $s_\downarrow(t) = \frac{1}{1+t}$ can be obtained as the output of Φ with $k = 1$ and $b(t) = t$. The desired functions can then be obtained by multiplying $s_\uparrow$ or $s_\downarrow$ with t.

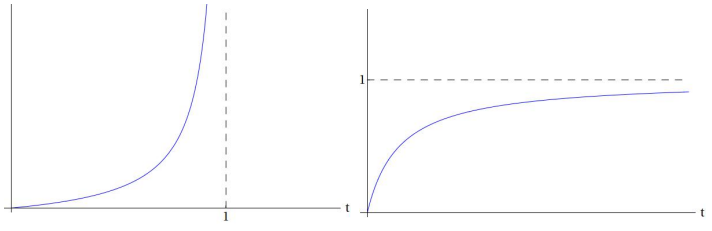


Figure 4.4: Plot of the functions $t \mapsto \frac{t}{1-t}$ (left) and $t \mapsto \frac{t}{1+t}$ (right).

□

Therefore, if we have a function $u(t)$ with a desired limit as $t \to \infty$, we can perform a composition of u with the infinite speedup to obtain the desired limit as $t \to 1^-$. The reverse case is also possible; that is, we can convert a limit as $t \to 1^-$ into a limit as $t \to +\infty$ via a composition with the infinite slowdown.

4.4 Pseudonorm effectiveness

Our construction of the limit module relies on the notion of effective limit, which is given by the metric associated to the underlying space $\mathcal{X}$. The advantage of this approach is that it requires only a minimal structure on $\mathcal{X}$ (complete metric space), and thus it can be applied quite generally. However, Chapters 2 and 3 provided evidence for the prevalence of Fréchet spaces in our research. Since the topology in these spaces is induced by a family of pseudonorms, we may desire to define a suitable notion of effective limits that takes this into consideration. Since a metric can be inferred from the pseudonorms (recall Proposition 2.2.13), we may expect some equivalence between both notions. In this section we formalize this argumentation.

Definition 4.4.1 (**Moduli of convergence for pseudonorms**).

1. A *discrete modulus of convergence for pseudonorms* is a function $N : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that for each $n \in \mathbb{N}$, $N(n, \cdot)$ is nondecreasing.

2. A *continuous modulus of convergence for pseudonorms* is a function $T : \mathbb{N} \to C(\mathbb{T}, \mathbb{R})$ such that for each $n \in \mathbb{N}$, $T(n) \in C(\mathbb{T}, \mathbb{R})$ is nonnegative and nondecreasing.

Observe that for each $n \in \mathbb{N}$, the n-section of a (discrete or continuous) modulus of convergence for pseudonorms is itself a (discrete or continuous) modulus of convergence in the underlying space.

Definition 4.4.2 (**Effective limits on Fréchet spaces**).

1. Let N be a discrete modulus of convergence for pseudonorms and $\{g_n\} \in \mathcal{X}^{\mathbb{N}}$. Then $\{g_n\}$ is an *N-Fréchet Cauchy sequence* (or an N-FC sequence) if

$$\text{for all } \nu \in \mathbb{N}, n \in \mathbb{N}, \text{ for all } j, k \in \mathbb{N} \text{ with } j, k \geq N(n, \nu) \text{ one has } \|g_j, g_k\|_n < 2^{-\nu}.$$

2. Let T be a continuous modulus of convergence for pseudonorms and $u \in C(\mathbb{T}, \mathcal{X})$. Then u is a *T-Fréchet Cauchy stream* (or a T-FC stream) if

$$\text{for all } \tau \in \mathbb{T}, n \in \mathbb{N}, \text{ for all } s, t \in \mathbb{T} \text{ with } s, t \geq T(n, \tau) \text{ one has } \|u(s), u(t)\|_n < 2^{-\tau}.$$

For the following lemma, we shall assume that the metric in $\mathcal{X}$ is induced by the pseudonorms as

$$d(u, v) = \sum_{n \in \mathbb{N}} w_n \gamma(\|u - v\|_n), \quad \text{with } w_n = 2^{-n} \text{ and } \gamma(t) = \min(t, 1), \tag{4.1}$$

which satisfy the assumptions in Proposition 2.2.13 (see also Proposition 2.2.15).

Lemma 4.4.3 (**Equivalence between effective limits**).

1. *Let N be a discrete modulus of convergence and $g \in \mathcal{X}^{\mathbb{N}}$ an N-convergent Cauchy sequence. Then g is an $\tilde{N}$-FC sequence, where $\tilde{N}(n, \nu) = N(n + \nu)$; moreover, if N is computable, so is $\tilde{N}$.*

2. *Let T be a continuous modulus of convergenge and $u \in C(\mathbb{T}, \mathcal{X})$ a T-convergent Cauchy stream. Then u is a $\tilde{T}$-FC stream, where $\tilde{T}(n, \tau) = T(n + \tau)$; moreover, if T is GPAC-generable, so is $\tilde{T}(n)$ for each n.*

3. *Let $\tilde{N}$ be a discrete modulus of convergence for pseudonorms and $g \in \mathcal{X}^{\mathbb{N}}$ an $\tilde{N}$-FC sequence. Then g is an N-convergent Cauchy sequence, where $N(\nu) = \max_{n \leq \nu+1} \tilde{N}(n, \nu+1)$; moreover, if $\tilde{N}$ is computable, so is N.*

4. *Let $\tilde{T}$ be a continuous modulus of convergence for pseudonorms and $u \in C(\mathbb{T}, \mathcal{X})$ a $\tilde{T}$-FC stream. Then u is a T-convergent Cauchy stream, where $T(\tau) = \max_{n \leq \tau+2} \tilde{T}(n, \tau+1)$.*

Proof. To prove claim 1, we first observe that for each n, the function $\tilde{N}(n, \cdot) : \nu \mapsto N(n+\nu)$ is nonnegative and nondecreasing (since N is nonnegative and nondecreasing), so that $\tilde{N}$ is a discrete modulus of convergence for pseudonorms. It is also clear from inspection that if N is computable, so is $\tilde{N}$.

Next, we take $\nu \in \mathbb{N}$, $n \in \mathbb{N}$ and $j, k \in \mathbb{N}$ with $j, k \geq \tilde{N}(n, \nu)$. By construction of $\tilde{N}$ this means that $j, k \geq N(n+\nu)$ and thus, since g is an N-convergent Cauchy sequence, it follows that $d(g_j, g_k) < 2^{-n-\nu}$. By looking only at the n-th term in the sum in (4.1), we conclude that $w_n \gamma(\|g_j - g_k\|_n) < 2^{-n-\nu}$, which implies that $\min(\|g_j - g_k\|_n, 1) < 2^{-\nu}$. Since $2^{-\nu} \leq 1$, we then have that $\|g_j - g_k\|_n < 2^{-\nu}$. Thus g is an $\tilde{N}$-FC sequence.

To prove claim 2, we first observe that for each n, the function $t \mapsto T(n+t)$ is nonnegative and nondecreasing (since T is nonnegative and nondecreasing), so that $\tilde{T}$ is a continuous modulus of convergence for pseudonorms. Moreover, each $t \mapsto T(n+t)$ is computable since it is the composition of T with the function $t \mapsto n+t$, which can be obtained using one constant and one adder module. As a side remark, the procedure that maps n into a GPAC $\mathcal{G}_n$ generating the corresponding $\tilde{T}(n)$ is also computable on n.

The remainder of the claim can be proved, *mutatis mutandis*, as in claim 1.

To prove claim 3, we first see that the function $\nu \mapsto \max_{n \leq \nu+1} \tilde{N}(n, \nu+1)$ is nonnegative and nondecreasing, since $\tilde{N}(n, \cdot)$ is nonnegative and nondecreasing for each n, so that N is a discrete modulus of convergence. It is also clear that if $\tilde{N}$ is computable, so is N, since taking maxima is a computable operation in $\mathbb{N}$.

Next, we take $\nu \in \mathbb{N}$ and $j, k \in \mathbb{N}$ with $j, k \geq N(\nu)$. By construction of N this means that $j, k \geq \tilde{N}(n, \nu+1)$ for all $n \leq \nu+1$ and thus, since g is an $\tilde{N}$-FC sequence, it follows that $\|g_j - g_k\|_n < 2^{-\nu-1}$ for all $n \leq \nu+1$. By splitting the sum in (4.1), we can see that

$$\begin{aligned} d(g_j, g_k) &= \sum_{n \in \mathbb{N}} w_n \gamma(\|g_j - g_k\|_n) \\ &\leq \sum_{1 \leq n \leq \nu+1} 2^{-n} \|g_j - g_k\|_n + \sum_{n > \nu+1} 2^{-n} \\ &< \sum_{1 \leq n \leq \nu+1} 2^{-n} 2^{-\nu-1} + \sum_{n > \nu+1} 2^{-n} \\ &\leq 2^{-\nu-1} + 2^{-\nu-1} = 2^{-\nu}, \end{aligned}$$

so that g is an N-convergent Cauchy sequence.

To prove claim 4, we first see that the function $t \mapsto \max_{n \leq \tau+2} \tilde{T}(n, \tau+1)$ is nonnegative and nondecreasing, so that T is a continuous module of convergence. The remainder of the claim can be proved, *mutatis mutandis*, as in claim 3. □

4.5 Computability of the Gamma function

Our motivation for considering limit operators is the computability of the gamma function,

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt,$$

which is not differentially algebraic (and thus, not Shannon GPAC-generable)[1]. There are known differential equations in two variables related to the gamma function; for example (see [OLBC10, p. 174]), if we define the incomplete gamma functions

$$\gamma_{i1}(t,x) = \int_0^t s^{x-1} e^{-s} ds; \tag{4.2}$$

$$\gamma_{i2}(t,x) = \int_t^\infty s^{x-1} e^{-s} ds, \tag{4.3}$$

then both incomplete gamma functions satisfy the differential equation (for $w = w(t,x)$)

$$\frac{d^2 w}{dt^2} + \left(1 + \frac{1-x}{t}\right) \frac{dw}{dt} = 0; \tag{4.4}$$

we shall now try to implement such relations on our analog networks.

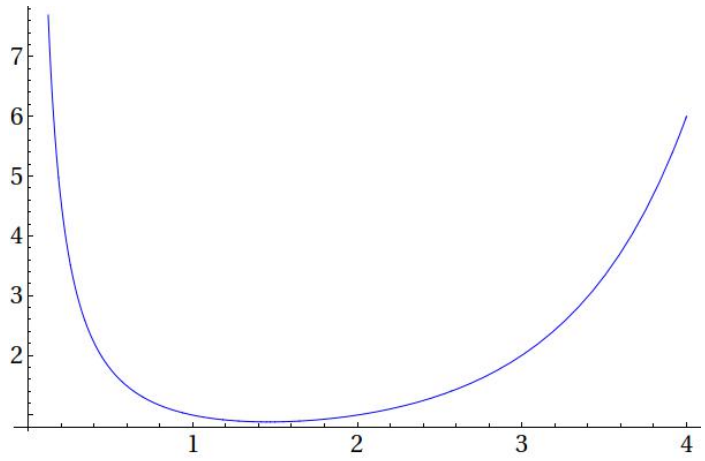


Figure 4.5: Plot of the gamma function.

Observe that the $\mathcal{X}$-GPAC includes a constant module for any function in $\mathcal{X}$ (by constant we mean: not dependent on the time variable t), and in particular it includes a constant module for the gamma function itself! Of course, this is not an interesting way to obtain the gamma function, as one could then raise the question 'how can we generate the constant modules?'

As a side note, we could answer this question by defining a notion of relative computability. For example, if we consider a subclass $\mathcal{G} \subseteq \mathcal{X}$ of 'admissible constants' we could say that 'f is $\mathcal{X}$-GPAC-generable relative to $\mathcal{G}$' if there is an $\mathcal{X}$-GPAC which generates f and whose constant inputs are all in $\mathcal{G}$. With this notion, we would obtain that Γ is $\mathcal{X}$-GPAC-generable relative to itself, but this is

[1] Proved in [Höl86], mentioned in [Sha41].

not the approach we are interested in.

We shall thus present a more elaborate construction; the main idea is to obtain the gamma function as the limit of a function in two variables,

$$\Gamma(x) = \lim_{t\to\infty} \gamma(t,x),$$

for some function $\gamma \in C(\mathbb{T}, \mathcal{X})$ which will be specified shortly. As remarked in Section 4.3, the choice of limit $t \to \infty$ is arbitrary, as we can take infinite speedups and consider, e.g., a limit $t \to 1^-$. Since $\Gamma(x)$ has a pole at $x = 0$, we need to consider a space where functions are defined in a region "away from" $x = 0$. For simplicity, we shall take $\mathcal{X} = C[1, +\infty)$. We also observe that (4.4) is undetermined at $t = 0$, and it would allow initial conditions $w|_{t=0} = \frac{dw}{dt}|_{t=0} = 0$, for which $w \equiv 0$ is a different solution. Since well-posedness is desired, we must avoid starting at $t = 0$; therefore, we consider integrals starting at $t = 1$, writing

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt = \int_0^1 t^{x-1}e^{-t}dt + \int_1^\infty t^{x-1}e^{-t}dt.$$

The next step is to apply a change of variables in order to obtain integrals of the form $\int_0^\infty$; to be precise, we apply $t \mapsto s = \frac{1-t}{t}$ on the first integral and $t \mapsto s = t - 1$ on the second integral, obtaining

$$\int_0^1 t^{x-1}e^{-t}dt = \int_0^\infty \left(\frac{1}{1+s}\right)^{x+1} e^{-1/(1+s)}ds = \lim_{t\to+\infty} \gamma_1(t,x);$$
$$\int_1^\infty t^{x-1}e^{-t}dt = \int_0^\infty (1+s)^{x-1}e^{-(1+s)}ds = \lim_{t\to+\infty} \gamma_2(t,x),$$

where

$$\gamma_1(t,x) = \int_0^t (1+s)^{-(x+1)}e^{-1/(1+s)}ds;$$
$$\gamma_2(t,x) = \int_0^t (1+s)^{x-1}e^{-(1+s)}ds.$$

We proceed to show that γ_1, γ_2 are $\mathcal{X}$-GPAC-generable.

Computation of γ_1: by taking derivatives in time, we see that

$$\frac{d\gamma_1}{dt} = (1+t)^{-(x+1)}e^{-1/(1+t)};$$
$$\frac{d^2\gamma_1}{dt^2} = -(x+1)(1+t)^{-(x+2)}e^{-1/(1+t)} + (1+t)^{-(x+3)}e^{-1/(1+t)} = -\frac{x+xt+t}{(1+t)^2}\frac{d\gamma_1}{dt}; \tag{4.5}$$

moreover, we have initial conditions

$$\gamma_1(0,x) = 0, \quad \frac{d\gamma_1}{dt}(0,x) = 1/e.$$

We can look at the PDE (4.5) as an ODE in t with a parameter x. It is then easy to check that it defines a well-posed problem since the multiplying factor $u_1(t,x) = -\frac{x+xt+t}{(1+t)^2}$ is defined for all $t \in \mathbb{T}$. As an intermediate step in generating γ_1 with an $\mathcal{X}$-GPAC, via (4.5), we generate the multiplying factor u_1, and to achieve this we consider the function $s_\downarrow(t) = \frac{1}{1+t}$, which is GPAC-generable by the proof of Proposition 4.3.1. We can thus construct $u_1 = -(x + xt + t)s_\downarrow^2$ and obtain γ_1 with an

$\mathcal{X}$-GPAC as in Figure 4.6, which implements (4.5).

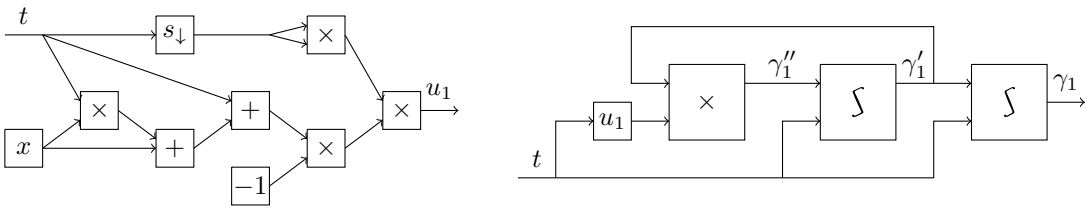


Figure 4.6: Construction of $u_1(t) = -\frac{x+xt+t}{(1+t)^2}$ and $\gamma_1(t,x)$.

Computation of γ_2: by taking derivatives in time, we see that

$$\begin{aligned}\frac{d\gamma_2}{dt} &= (1+t)^{x-1}e^{-(1+t)};\\ \frac{d^2\gamma_2}{dt^2} &= (x-1)(1+t)^{x-2}e^{-(1+t)} - (1+t)^{x-1}e^{-(1+t)} = \frac{x-t-2}{1+t}\frac{d\gamma_2}{dt};\end{aligned} \tag{4.6}$$

moreover, we have initial conditions

$$\gamma_2(0,x) = 0, \quad \frac{d\gamma_2}{dt}(0,x) = 1/e.$$

We can look at the PDE (4.6) as an ODE in t with a parameter x. It is then easy to check that it defines a well-posed problem, since the multiplying factor $u_2(t,x) = \frac{x-t-2}{1+t}$ is defined for all $t \in \mathbb{T}$. As an intermediate step in generating γ_2 with an $\mathcal{X}$-GPAC, via (4.6), we generate the multiplying factor u_2, and to achieve this we again consider the function $s_\downarrow(t) = \frac{1}{1+t}$ from proof of Proposition 4.3.1. We can thus construct $u_2 = (x-t-2)s_\downarrow$ and obtain γ_2 with an $\mathcal{X}$-GPAC as in Figure 4.7, which implements (4.6).

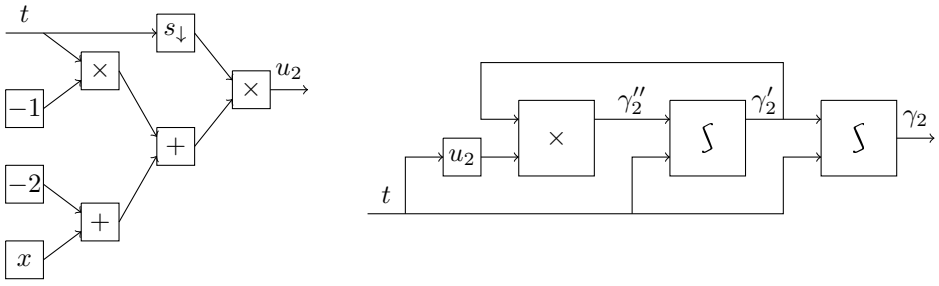


Figure 4.7: Construction of $u_2(t) = \frac{x-t-2}{1+t}$ and $\gamma_2(t,x)$.

Construction of Γ: We can obtain $\Gamma(x)$ as the limit

$$\Gamma(x) = \lim_{t\to\infty} \gamma_1(t,x) + \gamma_2(t,x),$$

which can in principle be obtained using a continuous limit module. However, we still need to determine the modulus of convergence of our approximation, which will be done with two technical lemmas.

Lemma 4.5.1. *Let $T \in \mathbb{T}$. For any $x \in [1, +\infty)$ and any $t_1, t_2 \geq T$ one has*

$$|\gamma_1(t_1, x) - \gamma_1(t_2, x)| \leq 1/T.$$

Proof. Under the assumptions of the lemma, we have

$$\begin{aligned} |\gamma_1(t_1, x) - \gamma_1(t_2, x)| &= \left| \int_{t_2}^{t_1} (1+s)^{-(x+1)} e^{-1/(1+s)} ds \right| < \int_T^{\infty} (1+s)^{-(x+1)} e^{-1/(1+s)} ds \\ &< \int_T^{\infty} (1+s)^{-2} ds = \frac{1}{1+T} < \frac{1}{T}. \end{aligned}$$

□

Lemma 4.5.2. *Let $T \in \mathbb{T}$, and $k \in \mathbb{N}$. For any $x \in [1, k+1]$ and any $t_1, t_2 \geq T$ one has*

$$|\gamma_2(t_1, x) - \gamma_2(t_2, x)| \leq (k+1)!(T+1)^k e^{-(T+1)}.$$

Proof. Under the assumptions of the lemma, we have

$$\begin{aligned} |\gamma_2(t_1, x) - \gamma_2(t_2, x)| &= \left| \int_{t_2}^{t_1} (1+s)^{x-1} e^{-(1+s)} ds \right| < \int_T^{\infty} (1+s)^{x-1} e^{-(1+s)} ds \\ &\leq \int_T^{\infty} (1+s)^k e^{-(1+s)} ds = \int_{T+1}^{\infty} s^k e^{-s} ds < (k+1)!(T+1)^k e^{-(T+1)}, \end{aligned}$$

where the last inequality can be proved by induction on $k \in \mathbb{N}$. □

Therefore, the limits in γ_1, γ_2 become effective for suitable moduli of convergence. We can merge these two results and prove effectiveness of our construction, as in the next result. We continue to use the metric given by (4.1). Recall that $\mathcal{X} = C([1, \infty))$ is a Fréchet space with pseudonorms $\|g\|_n = \sup_{1 \leq x \leq n} |g(x)|$.

Lemma 4.5.3. *Let $\gamma = \gamma_1 + \gamma_2$, where γ_1, γ_2 are defined as in (4.5), (4.6). Then $\lim_{t \to \infty} \gamma(t) = \Gamma$; moreover, γ is a T-convergent Cauchy stream for $T(\tau) = C2^\tau$ with a suitably large constant C.*

Proof. Only the effectiveness of the limit remains to be proven. Let $\tau \in \mathbb{T}$, $T = T(\tau) = C2^\tau$ and take $t_1, t_2 \in \mathbb{T}$ with $t_1, t_2 \geq T$. We can write $d(\gamma(t_1), \gamma(t_2)) \leq d(\gamma_1(t_1), \gamma_1(t_2)) + d(\gamma_2(t_1), \gamma_2(t_2))$ and thus we can treat γ_1 and γ_2 separately.

To deal with γ_1, we use Lemma 4.5.1 to conclude that, for any $n \in \mathbb{N}^+$, we have

$$\|\gamma_1(t_1) - \gamma_2(t_2)\|_n \leq \frac{1}{T} \leq \frac{1}{C} 2^{-\tau};$$

thus, we obtain the bound

$$\begin{aligned} d(\gamma_1(t_1), \gamma_1(t_2)) &= \sum_{n=1}^{\infty} 2^{-n} \min\{1, \|\gamma_1(t_1) - \gamma_1(t_2)\|_n\} \\ &\leq \sum_{n=1}^{\infty} 2^{-n} \|\gamma_1(t_1) - \gamma_1(t_2)\|_n \leq \sum_{n=1}^{\infty} 2^{-n} \frac{1}{C} 2^{-\tau} = \frac{1}{C} 2^{-\tau}, \end{aligned}$$

which is smaller than $2^{-\tau-1}$ for a suitably large C (namely, for $C > 2$).

To deal with γ_2, we use Lemma 4.5.2 to conclude that, for any $n \in \mathbb{N}^+$, we have

$$\|\gamma_2(t_1) - \gamma_2(t_2)\|_n \leq n!(T+1)^{n-1}e^{-(T+1)};$$

next, we shall take $N = \lceil \tau \rceil + 2$, so that $\tau + 2 \leq N < \tau + 3$. By splitting the sum, we obtain the bound

$$\begin{aligned}
d(\gamma_2(t_1), \gamma_2(t_2)) &= \sum_{n=1}^{\infty} 2^{-n} \min\{1, \|\gamma_1(t_1) - \gamma_1(t_2)\|_n\} && \text{(4.7a)}\\
&\leq \sum_{n=1}^{N} 2^{-n}\|\gamma_1(t_1) - \gamma_1(t_2)\|_n + \sum_{n=N+1}^{\infty} 2^{-n} && \text{(4.7b)}\\
&\leq \sum_{n=1}^{N} 2^{-n} n!(T+1)^{n-1}e^{-(T+1)} + 2^{-N} && \text{(4.7c)}\\
&\leq N!(T+1)^{N-1}e^{-(T+1)} \sum_{n=1}^{N} 2^{-n} + 2^{-\tau-2} && \text{(4.7d)}\\
&< eN^{N+1/2}e^{-N}(T+1)^{N-1}e^{-(T+1)} + 2^{-\tau-2} && \text{(4.7e)}\\
&= \exp\{(N+1/2)\log(N) - N + (N-1)\log(T+1) - T\} + 2^{-\tau-2} && \text{(4.7f)}\\
&< \exp\{(\tau+7/2)\log(\tau+3) - \tau - 2 + (\tau+2)\log(C2^{\tau}+1) - C2^{\tau}\} + 2^{-\tau-2}, && \text{(4.7g)}
\end{aligned}$$

where (4.7e) is justified by Stirling's approximation (2.53). The last step can be further taken to be smaller than $\exp\{-\log(2)(\tau+2)\} + 2^{-\tau-2} = 2^{-\tau-1}$ for a suitably large C that does not depend on τ (because the term $C2^{\tau}$ in the exponential largely dominates all other terms; numerically, we have found that $C > 2.85216$ suffices).

Combining the two bounds, we conclude that $d(\gamma(t_1), \gamma(t_2)) < 2^{-\tau}$ and therefore γ is a T-convergent Cauchy stream. □

Theorem 20. *The gamma function is LGPAC-generable.*

Proof. By Lemma 4.5.3, the gamma function Γ can be seen as the T-convergent limit of some function γ. Moreover, by the preceding discussion, both γ and T can be seen to be $\mathcal{X}$-GPAC-generable; in particular, γ is the sum of two $\mathcal{X}$-GPAC-generable functions. Thus we can devise an LGPAC that generates Γ, as in Figure 4.8. □

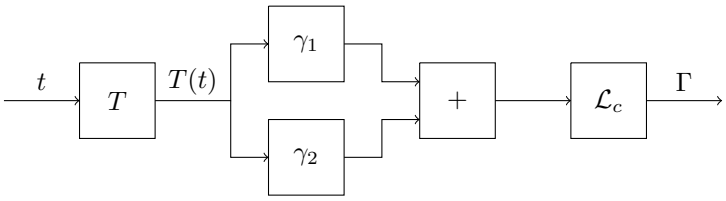


Figure 4.8: Construction of the gamma function; T denotes an exponential speedup $T(\tau) = C2^{\tau}$.

4.6 Computability of the Riemann zeta function

Our next case study concerns the computation of the Riemann zeta function, which for complex numbers with real part greater than 1 is given by

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}. \tag{4.8}$$

This function has a pole at $z = 1$ and thus we should consider a space of functions defined in a region "away from" $z = 1$. In particular, we take $\mathcal{X} = C[2, \infty)$; in other words, we shall be interested in computing $\zeta(x)$ for real values of x larger or equal to 2.

We need a representation of the Riemann zeta function that is amenable to our framework of analog networks. Fortunately, there are known integral representations that we can use, such as

$$\zeta(x) = \frac{1}{\Gamma(x)} \int_0^{\infty} \frac{t^{x-1}}{e^t - 1} dt, \tag{4.9}$$

or the Abel-Plana formula [Abe65, Pla20]

$$\zeta(x) = \frac{2^x}{x-1} - 2^x \int_0^{\infty} \frac{\sin(x \arctan t)}{(1+t^2)^{x/2}(e^{\pi t+1})} dt. \tag{4.10}$$

The latter formula will allow us to express the zeta function as the limit of a function in two variables,

$$\zeta(x) = \lim_{t \to \infty} \zeta_1(t, x),$$

for a function ζ_1 which computes the bounded integral

$$\zeta_1(t, x) = \frac{2^x}{x-1} - 2^x \int_0^{t} \frac{\sin(x \arctan s)}{(1+s^2)^{x/2}(e^{\pi s+1})} ds. \tag{4.11}$$

For such a function, we have $\zeta_1(0, x) = \frac{2^x}{x-1}$ and $\frac{d\zeta_1}{dt} = -2^x \zeta_2$, where

$$\zeta_2(t, x) = \frac{\sin(x \arctan t)}{(1+t^2)^{x/2}(e^{\pi t+1})}. \tag{4.12}$$

Lemma 4.6.1. *The function ζ_2 defined in* (4.12) *is $\mathcal{X}$-GPAC-generable.*

Proof. This requires several steps, so we just provide a sketch of the construction:

1. the function $t \mapsto \frac{1}{1+t^2}$ is GPAC-generable; it can be given as the output of the inverter (from Example 3.8.4) with inputs $k = 1$ and $b(t) = t^2$;
2. the function $t \mapsto \arctan t$ is GPAC-generable; observe that $(\arctan t)' = \frac{1}{1+t^2}$ and use step 1;
3. the function $(t, x) \mapsto \sin(x \arctan t)$ is $\mathcal{X}$-GPAC-generable; compose $(t, x) \mapsto x \arctan t$ (from step 2) with $t \mapsto \sin(t)$;
4. the function $(t, x) \mapsto (1 + t^2)^{-x/2}$ is $\mathcal{X}$-GPAC-generable; if $u(x, t) = (1 + t^2)^{-x/2}$ then $\frac{du}{dt} = -\frac{xt}{1+t^2} u$, with $u(0, x) = 1$; use step 1;
5. the function $t \mapsto e^{-\pi t-1}$ is GPAC-generable; compose $t \mapsto -\pi t - 1$ with $t \mapsto e^t$;

6. the function ζ_2 is $\mathcal{X}$-GPAC-generable; write $\zeta_2(t,x) = \sin(x \arctan t)(1+t^2)^{-x/2}e^{-\pi t - 1}$ and use steps 3, 4, 5.

□

Theorem 21. *The Riemann zeta function is LGPAC-generable.*

Proof. We can obtain ζ_1 (from (4.11)) by feeding ζ_2 (which is $\mathcal{X}$-GPAC-generable by Lemma 4.6.1) into an integrator module and using constants $\frac{2^x}{x-1}$, -2^x. We can obtain the Riemann zeta function by feeding ζ_1 into an effective limit module. Thus we can devise an LGPAC that generates ζ, as in Figure 4.9.

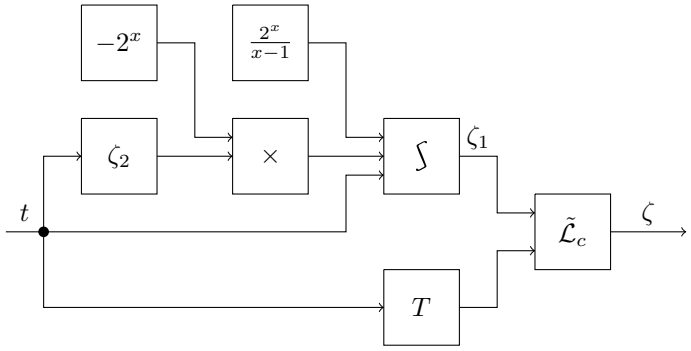


Figure 4.9: Construction of the Riemann zeta function; T denotes a suitable continuous modulus of convergence.

The only thing left is to prove the effectiveness of the convergence. In order to do that we shall prove that a linear modulus of convergence $T(\tau) = C\tau$, for a suitable large constant C, is sufficient. The following calculations are similar to those done for Lemmas 4.5.1, 4.5.2 and 4.5.3. To start, we recall that $\mathcal{X} = C[2,\infty)$ is a Fréchet space with pseudonorms $\|g\|_n = \sup_{2\le x\le n} |g(x)|$. Let $T \in \mathbb{T}$, $k \in \mathbb{N}$ with $k \ge 2$, $x \in [2,k]$ and $t_1, t_2 \in \mathbb{T}$ with $t_1, t_2 \ge T$; then we have the bound

$$\begin{aligned}
|\zeta_1(t_1,x) - \zeta_1(t_2,x)| &= \left| 2^x \int_{t_2}^{t_1} \frac{\sin(x\arctan t)}{(1+t^2)^{x/2}e^{\pi t+1}} dt \right| \\
&< 2^x \int_T^\infty \left| \frac{\sin(x\arctan t)}{(1+t^2)^{x/2}e^{\pi t+1}} \right| dt \\
&\le 2^k \int_T^\infty \frac{1}{e^{\pi t+1}} dt = \frac{2^k}{e\pi} e^{-\pi T}.
\end{aligned}$$

Thus, for any $k \ge 2$ and any $t_1, t_2 \ge T(\tau)$ we have

$$\|\zeta_1(t_1) - \zeta_1(t_2)\|_k \le \frac{2^k}{e\pi} e^{-\pi T} = \frac{2^k}{e\pi} e^{-\pi C\tau}. \tag{4.13}$$

Next, let us take $N = \lceil \tau \rceil + 1$, so that $\tau + 1 \le N < \tau + 2$. By splitting the sum, we obtain

$$d(\zeta_1(t_1), \zeta_1(t_2)) = \sum_{n=2}^{\infty} 2^{-n} \min(\|\zeta_1(t_1) - \zeta_1(t_2)\|_n, 1) \tag{4.14a}$$

$$\leq \sum_{n=2}^{N} 2^{-n} \|\zeta_1(t_1) - \zeta_1(t_2)\|_n + \sum n = N + 1^{\infty} 2^{-n} \tag{4.14b}$$

$$\leq \sum_{n=2}^{N} 2^{-n} \frac{2^n}{e\pi} e^{-\pi C\tau} + 2^{-N} \tag{4.14c}$$

$$\leq \frac{N-1}{e\pi} e^{-\pi C\tau} + 2^{-\tau-1} \tag{4.14d}$$

$$< \frac{\tau+1}{e\pi} e^{-\pi C\tau} + 2^{-\tau-1} \tag{4.14e}$$

$$= \exp\{-\pi C\tau + \log(\tau+1) - \log(e\pi)\} + 2^{-\tau-1}, \tag{4.14f}$$

where (4.14c) is justified by (4.13). Finally, the last step can be further taken to be smaller than $\exp\{-\log(2)(\tau+1)\} + 2^{-\tau-1} = 2^{-\tau}$ for a suitably large C that does not depend on τ (because the term $\pi C\tau$ dominates all other terms; numerically, we have found that $C > 0.25079$ suffices). Thus $d(\zeta_1(t_1), \zeta_1(t_2)) < 2^{-\tau}$, so that ζ_1 is a T-convergent Cauchy stream. Incidentally, since the lower bound for C is less than 1, the stream ζ_1 is id-convergent. □

4.7 Discussion

In this chapter we took a different direction from Chapter 3 and introduced limit modules to our model, arriving at a generalization which we called LGPAC. The main motivation was to prove that some non-differentially algebraic functions such as the gamma function can be generated in this framework. In some sense, that result was obtained before (see [Gra04]) by changing the notion of GPAC-generability to allow for approximability of functions.

The idea of *approximability* is a cornerstone in many models of computability on continuous spaces, especially those that use classically computable functions (i.e. computable functions on the naturals) as a starting point. This is a consequence of the fact that many continuous spaces are typically represented using a dense countable subset and codes of convergent sequences (as we shall explain in the next chapter). Then, to say that a function is computable is to assert that its values can be obtained up to a prescribed precision in an effective way.

Chapter 5

Tracking computability of GPAC-generable functions

The goal of this chapter is to connect the model of computation presented in this thesis (the $\mathcal{X}$-GPAC and LGPAC) with other models of computability in continuous spaces. Namely, we shall look into the notion of tracking computability presented in [TZ04] and show that, under some suitable conditions, the functions generated by a GPAC are tracking computable. We will consider a version of the GPAC which combines the constructions presented in Chapters 3 and 4; namely, it will have both a differential module (as in Definition 3.4.2) and a (one-input, continuous) limit module (as in Definition 4.2.4). As a technical note, we will have to slightly adapt our notions of induced operator and GPAC semantics in order to prove computability of the desired functionals.

We begin by introducing the notions of computable structures and tracking computable functions studied by Tucker and Zucker, providing examples for the spaces of interest in our GPAC model. We then prove tracking computability of the functions associated to the LGPAC modules (including the differential and continuous limit modules) and of the induced operator of an LGPAC. Finally, we attempt to prove tracking computability of LGPAC-generable functions; in order to achive this, we assume an additional condition on our LGPAC, which we call *effective local reversibility*.

In regards to the original content of this chapter, we remark that the paper [BCGH07] already shows an equivalence between a GPAC model (which includes approximability) and computable analysis, which is likely equivalent to tracking computability (papers [TZ04] and [TZ05] have results in the direction of a formal proof). The main results obtained in this chapter (namely Lemma 5.2.1 and Theorems 22 and 23) can be seen as new insofar as they are expressed in a new framework, where approximability is incorporated on the GPAC model by means of limit modules (as we argued in Chapter 4). The technical notion of effective local reversibility (Definition 5.3.4) is also an original idea.

5.1 Computable structures and tracking computable functions

The procedure for defining tracking computability in general spaces has been extensively documented by many authors. The basic construction consists in taking an enumeration of a countable dense subset, defining computable elements as those given by effective Cauchy sequences, and considering tracking functions.

As usual, we consider complete metric spaces $\mathcal{X}$. In this chapter, we must include the extra assumption that $\mathcal{X}$ be *separable* (i.e. has a countable dense subset).

Definition 5.1.1 (**Enumeration**)**.** Let $\mathcal{X}$ be a separable, complete metric space and $\mathcal{X}_c$ a countable, dense subset of $\mathcal{X}$. An *enumeration* of $\mathcal{X}_c$ is a surjective total function $\alpha : \mathbb{N} \to \mathcal{X}_c$.

We briefly recall the notion of *effective Cauchy sequence* given in Definition 4.2.3. For a complete metric space $\mathcal{X}$ and a sequence (x_n) in $\mathcal{X}$, we say that (x_n) is an effective Cauchy sequence if there exists a function $N : \mathbb{N} \to \mathbb{N}$ such that

- N is nondecreasing and computable (in the traditional sense);
- for all $\nu \in \mathbb{N}$, for all $m, n \in \mathbb{N}$ with $m, n \geq N(\nu)$ one has $d(x_m, x_n) < 2^{-\nu}$.

The function N in the previous definition is called an *effective discrete modulus of convergence* (cf. Definition 4.2.1 and Remark 4.2.2).

In the next definition we fix a family of computable bijections $\langle \cdot, \cdots, \cdot \rangle : \mathbb{N}^M \to \mathbb{N}$, for $M \in \mathbb{N}^+$ (say, the Cantor pairing function $\langle \cdot, \cdot \rangle$ for $M = 2$ and its generalizations to higher dimensions). We also consider an enumeration $\{\cdot\} : \mathbb{N} \to (\mathbb{N} \rightharpoonup \mathbb{N})$ of the recursive functions (say, for $e \in \mathbb{N}$ the encoding of a one-input, one-output Turing machine, $\{e\}$ is the corresponding recursive function).

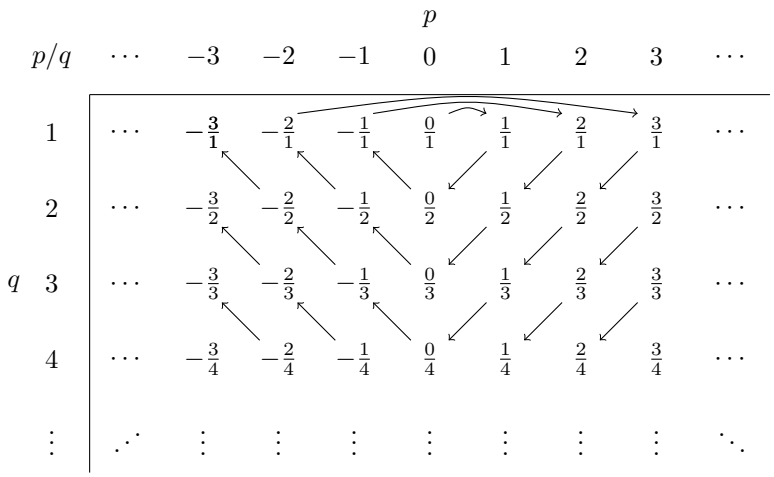
Definition 5.1.2 (**Computability structure**)**.** Let $\mathcal{X}$ be a complete metric space, $\mathcal{X}_c$ a countable, dense subset of $\mathcal{X}$ and α an enumeration of $\mathcal{X}_c$. We define a *computability structure* $(\Omega_{\bar{\alpha}}, C_{\bar{\alpha}}, \bar{\alpha})$ as follows.

1. The set of *valid codes* $\Omega_{\bar{\alpha}}$, is the subset of $\mathbb{N}$ given by encodings of pairs of numbers $c = \langle e, m \rangle$ such that e is the index for a total recursive function $\{e\} : \mathbb{N} \to \mathbb{N}$, m is the index for an effective discrete modulus of convergence $\{m\} : \mathbb{N} \to \mathbb{N}$ and the sequence $(\alpha(\{e\}(n)))$ is effective Cauchy with modulus of convergence $\{m\}$.
2. The set of *computable elements* $C_{\bar{\alpha}}$ is the subset of $\mathcal{X}$ consisting of those $x \in \mathcal{X}$ for which there exists a valid code $c = \langle e, m \rangle \in \Omega_{\bar{\alpha}}$ such that $x = \lim \alpha(\{e\}(n))$.
3. The *partial enumeration* $\bar{\alpha} : \mathbb{N} \rightharpoonup \mathcal{X}$ is the function with domain $\Omega_{\bar{\alpha}}$, range $C_{\bar{\alpha}}$, and such that for any $c = \langle e, m \rangle \in \Omega_{\bar{\alpha}}$, $\bar{\alpha}(c) = \lim \alpha(\{e\}(n))$.

Example 5.1.3 (**Computability on** $\mathbb{R}$)**.** Let us construct a computability structure on the space $\mathcal{X} = \mathbb{R}$. In this case, we take the rationals as our countable dense subset, $\mathcal{X}_c = \mathbb{Q}$. We take $\alpha = \alpha_{\mathbb{Q}}$ to be any 'easily definable' enumeration of the rationals (for example, by using the canonical enumeration on Figure 5.1). In this case, the set of computable elements $C_{\bar{\alpha}_{\mathbb{Q}}}$ coincides with the familiar set of computable real numbers (also called recursive reals or constructible reals), as in [Tur36].

Example 5.1.4 (**Computability on** $C(\mathbb{R})$)**.** Let us construct a computability structure on the space $\mathcal{X} = C(\mathbb{R})$ of continuous real functions. In this case, we take $\mathcal{X}_c$ to be a countable subset of piecewise linear rational functions, which is defined as follows. For each $N \in \mathbb{N}$ and each tuple $(p_{-N^2}, \ldots, p_{-1}, p_0, p_1, \ldots, p_{N^2})$ of $2N^2 + 1$ rational numbers, we can consider a function $f : \mathbb{R} \to \mathbb{R}$ such that

- $f(x) = p_{-N^2}$ for every $x \leq -N$ and $f(x) = p_{N^2}$ for every $x \geq N$;
- $f(j/N) = p_j$ for each $j = -N^2, \ldots, 0, \ldots, N^2$;



n	p	q	$\alpha(n)$
0	0	1	0
1	1	1	1
2	0	2	0
3	−1	1	−1
4	2	1	2
5	1	2	1/2
6	0	3	0
7	−1	2	−1/2
8	−2	1	−2
9	3	1	3

Figure 5.1: Enumeration of the rationals; we have that $\alpha(n) = \frac{p}{q}$.

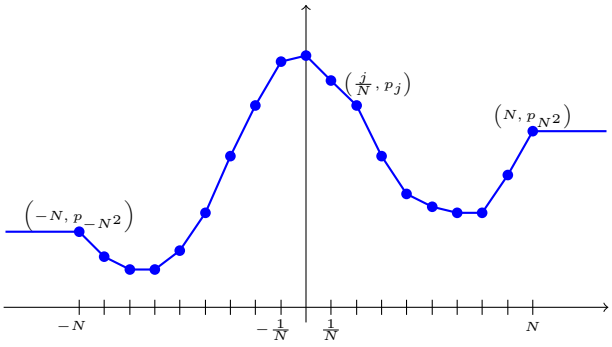


Figure 5.2: A piecewise linear rational function.

- f is piecewise linear on the interval $[j/N, (j+1)/N]$, for each $j = -N^2, \ldots, 0, \ldots, N^2 - 1$.

In this way, the role of N is both to increase the 'window size' and decrease the 'step size' of our approximation (see Figure 5.2). By using the bijections of type $\mathbb{N}^2 \to \mathbb{N}$ and $\mathbb{N}^{2N^2+1} \to \mathbb{N}$, and the enumeration $\alpha_{\mathbb{Q}}$ from the previous example, we can define an enumeration $\alpha_{\mathcal{X}} : \mathbb{N} \to \mathcal{X}_c$. In particular, the enumeration is as follows: for $n = \langle N, \langle m_{-N^2}, \ldots, m_{N^2} \rangle \rangle$, we define $\alpha_{\mathcal{X}}(n)$ to be the stream u built from N and the tuple $(p_{-N^2}, \ldots, p_{N^2})$ where $p_j = \alpha_{\mathbb{Q}}(m_j)$ for each $j = -N^2, \ldots, 0, \ldots, N^2)$.

We must briefly comment on the requirement that $\mathcal{X}_c$ be dense in $C(\mathbb{R})$. Recall that the metric (and the topology) of interest in $C(\mathbb{R})$ is induced by the family of pseudonorms $\|f\|_N = \sup_{-N \leq x \leq N} |f(x)|$. Since any continuous function can be approximated on any compact set by one of these piecewise linear rational functions, it follows that for any continuous function f we can devise a sequence f_N of piecewise linear rational functions such that $f_N \to f$ in the underlying topology.

Finally, we can apply the construction of Definition 5.1.2 and consider the set of computable elements $C_{\bar{\alpha}}$. In this case, this set coincides with the familiar set of computable real functions, as seen in [PER89, Wei00], among others.

Example 5.1.5 (Computability on $C^1(\mathbb{T},\mathcal{X})$). Given a computability structure on a Fréchet space $\mathcal{X}$ we shall construct a computability structure on the space of $\mathcal{X}$-streams $\mathcal{Z} = C^1(\mathbb{T},\mathcal{X})$. We shall consider the case $\mathbb{T} = [0,1]$ (bounded time); the case $\mathbb{T} = [0,T]$ is dealt in the exact same manner and the case $\mathbb{T} = [0,\infty)$ requires the additional trick presented in Example 5.1.4 of using a natural N for both 'window size' and 'step size'.

Let $(\mathcal{X}_c, \alpha_{\mathcal{X}})$ be an enumerated countable dense subset. We wish to apply the same principle as in Example 5.1.4, that is, construct an interpolant from a finite amount of 'data points'. However, in this case we cannot use piecewise linear functions, because we need to work with continuously differentiable functions. Recall that a function $u : \mathbb{T} \to \mathcal{X}$ is in $C^1(\mathbb{T},\mathcal{X})$ if the expression

$$v(t) = \lim_{h\to 0} \frac{u(t+h) - u(t)}{h}$$

is well-defined for all $t \in \mathbb{T}$ and defines a continuous function of time (that is, $v \in C(\mathbb{T},\mathcal{X})$). Therefore, the idea of our construction is to approximate v by a piecewise linear function and then integrate the approximation with respect to the time variable.

Formally, for each $N \in \mathbb{N}$ and each tuple $(x_0, y_0, \dots, y_N)$ of $N+2$ elements in $\mathcal{X}_c$, we consider the functions $u, v : \mathbb{T} \to \mathcal{X}$ such that

- $v(j/N) = y_j$ for each $j = 0, \dots, N$;
- v is piecewise linear on the interval $[j/N, (j+1)/N]$, for each $j = 0, \dots, N-1$;
- $u(t) = x_0 + \int_0^t v(s)ds$.

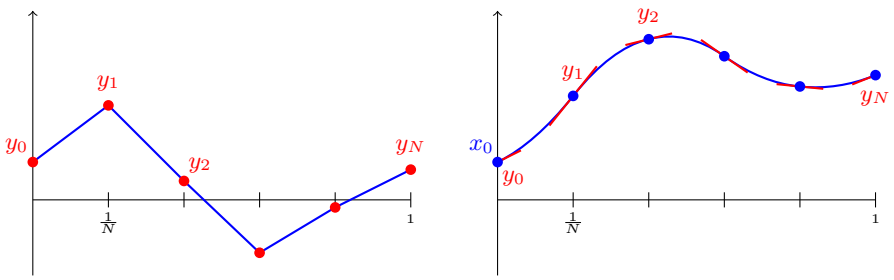


Figure 5.3: A continuous piecewise linear function v (left) and its integral, a C^1 piecewise quadratic function u (right). The data consist of an initial value x_0 and derivative values $y_0, \dots, y_N$ at collocated points.

We observe that, by construction, each u is continuously differentiable and piecewise quadratic (Figure 5.3). Now let $\mathcal{Z}_c$ be the space of functions u considered above, so that $\mathcal{Z}_c \subseteq \mathcal{Z}$. By using the bijections of type $\mathbb{N}^2 \to \mathbb{N}$ and $\mathbb{N}^{N+2} \to \mathbb{N}$, and the enumeration $\alpha_{\mathcal{X}}$, we can define an enumeration $\alpha_{\mathcal{Z}} : \mathbb{N} \to \mathcal{Z}_c$. Specifically: for $n = \langle N, \langle m_0, m'_0, \dots, m'_N\rangle\rangle$, we can define $\alpha_{\mathcal{Z}}(n)$ to be the stream u built from N and the tuple $(x_0, y_0, \dots, y_N)$ where $x_0 = \alpha_{\mathcal{X}}(m_0)$ and $y_j = \alpha_{\mathcal{X}}(m'_j)$ for each $j = 0, \dots N$.

Since $\mathcal{X}_c$ is countable, it is clear that $\mathcal{Z}_c$ is also countable. The fact that $\mathcal{Z}_c$ is dense in $\mathcal{Z}$ follows from basic topological principles (the proof is deferred to Proposition 5.1.6, after this example). Thus, we can apply the construction of Definition 5.1.2 and obtain the computability structure $(\Omega_{\bar{\alpha}_{\mathcal{Z}}}, C_{\bar{\alpha}_{\mathcal{Z}}}, \bar{\alpha}_{\mathcal{Z}})$.

We briefly remark that other constructible approaches are possible. For example, we could choose to approximate $u \in \mathcal{Z}$ by polynomials (with respect to the time derivative)

$$u(t) \approx x_0 + x_1 t + x_2 t^2 + \ldots + x_N t^N, \quad x_0, x_1, \ldots, x_N \in \mathcal{X}_c,$$

so that the data would consist of tuples $(x_0, \ldots, x_N)$. We could also consider cubic splines, so that $\mathcal{Z}_c$ is the space of C^1-functions which are piecewise cubic on each interval $[t_j, t_{j+1}]$. In this case, the data would consist of tuples $(x_0, \ldots, x_N, y_0, \ldots, y_N)$ corresponding to the values of a function and its derivative at the grid points t_j. Both these approaches can be seen to be equivalent to the first one.

Proposition 5.1.6. *The space $\mathcal{Z}_c$ considered in Example 5.1.5 is dense in $\mathcal{Z} = C^1(\mathbb{T}, \mathcal{X})$.*

Proof. Let $\| \cdot \|_n$ be the family of pseudonorms on $\mathcal{X}$ and consider the induced pseudonorms on $\mathcal{Z}$

$$\|u\|_n = \sup_{t \in \mathbb{T}} \|u(t)\|_n + \sup_{t \in \mathbb{T}} \|u'(t)\|_n.$$

We shall make the usual assumption that the metrics in $\mathcal{X}$ and $\mathcal{Z}$ are induced by the pseudonorms as (cf. Proposition 2.2.13, (2.16) and (4.1))

$$d(u_1, u_2) = \sum_{n=1}^{\infty} 2^{-n} \min(\|u_1 - u_2\|_n, 1). \tag{5.1}$$

Given $u \in \mathcal{Z}$, $\epsilon > 0$ and $M \in \mathbb{N}^+$, we shall construct an element $\tilde{u} \in \mathcal{Z}_c$ such that $\|u - \tilde{u}\|_n < \epsilon$ for $n = 1, \ldots, M$. This will imply that $\mathcal{Z}_c$ is dense in $\mathcal{Z}$, since we can infer a bound on $d_{\mathcal{Z}}(u, \tilde{u})$ from a bound on the pseudonorms, as per Proposition 2.2.15.

Let $v = u' \in C(\mathbb{T}, \mathcal{X})$. As $\mathbb{T} = [0, T]$ is compact and v is continuous, it follows that v is uniformly continuous, so that

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \forall t, s \in \mathbb{T} : \quad |t - s| < \delta \Rightarrow d_{\mathcal{X}}(v(t), v(s)) < \epsilon; \tag{5.2}$$

with a bit of effort (cf. Proposition 2.2.15), this property can be reexpressed in terms of the pseudonorms as

$$\forall \epsilon > 0 \quad \forall M \in \mathbb{N} \quad \exists \delta > 0 \quad \forall t, s \in \mathbb{T} \quad \forall n \leq M : \quad |t - s| < \delta \Rightarrow \|v(t) - v(s)\|_n < \epsilon. \tag{5.3}$$

Therefore, given $\epsilon > 0$ and $M \in \mathbb{N}$, we apply (5.3) and take $\delta > 0$ such that

$$\text{for any } n = 1, \ldots, M \text{ and any } t, s \in \mathbb{T} \text{ with } |t - s| \leq \delta, \text{ we have } \|v(t) - v(s)\|_n < \epsilon/6. \tag{5.4}$$

Now take $N \in \mathbb{N}$ such that $1/N < \delta$ (this N will be used for the 'step size' of our approximation) and take $x_0, y_0, \ldots, y_N \in \mathcal{X}_c$ which approximate $u(0), v(0), \ldots, v(1)$. In particular, we require that

$$\text{for any } n = 1, \ldots, M \text{ and any } j = 0, \ldots, N, \text{ we have } \|x_0 - u(0)\|_n, \|y_j - v(j/N)\|_n < \epsilon/6; \tag{5.5}$$

this can be achieved since $\mathcal{X}_c$ is dense in $\mathcal{X}$. Now we take $\tilde{u}, \tilde{v}$ as the functions constructed in Example 5.1.5 for the tuple $(x_0, y_0, \ldots, y_N)$. We shall show that $\tilde{u}$ is the desired approximation of u.

Fix some $t \in \mathbb{T}$ and some pseudonorm $\| \cdot \|_n$, $1 \leq n \leq M$. Let $0 \leq j \leq N$ be such that $t \in [j/N, (j+1)/N]$. Since $\tilde{v}$ is linear in that interval, we can write $\tilde{v}(t) = y_j + (Nt - j)(y_{j+1} - y_j)$.

Therefore we have the bound

$$\begin{aligned}\|\tilde{v}(t) - y_j\|_n &= \|(Nt - j)(y_{j+1} - y_j)\| \leq \|y_{j+1} - y_j\|_n \\ &\leq \|y_{j+1} - v((j+1)/N)\|_n + \|v((j+1)/N) - v(j/N)\|_n + \|v(j/N) - y_j\|_n \\ &< \epsilon/6 + \epsilon/6 + \epsilon/6 = \epsilon/2,\end{aligned}$$

where the last inequality is justified by (5.4) and (5.5). We can use this to obtain another bound,

$$\begin{aligned}\|v(t) - \tilde{v}(t)\|_n &\leq \|v(t) - v(j/N)\|_n + \|v(j/N) - y_j\|_n + \|y_j - \tilde{v}(t)\|_n \\ &< \epsilon/6 + \epsilon/6 + \epsilon/2 = 5\epsilon/6,\end{aligned}$$

where again the last inequality is justified by (5.4) and (5.5). Since $t \in \mathbb{T}$ was arbitrary, we can further conclude that $\|v - \tilde{v}\|_n < 5\epsilon/6$. We remark that, at this point, we have proven that the class of piecewise linear functions with endpoints in $\mathcal{X}_c$ is a dense subset of $C(\mathbb{T}, \mathcal{X})$. (To move from bounds in the pseudonorms to bounds in the metric, see Proposition 2.2.15).

To get a bound on the approximation of u, we let $t \in \mathbb{T}$ and notice that

$$\begin{aligned}\|u(t) - \tilde{u}(t)\|_n &= \left\| u(0) + \int_0^t v(s)ds - x_0 + \int_0^t \tilde{v}(s)ds \right\|_n \\ &\leq \|u(0) - x_0\|_n + \int_0^t \|v(s) - \tilde{v}(s)\|_n ds \\ &\leq \|u(0) - x_0\|_n + \|v - \tilde{v}\|_n \\ &< \epsilon/6 + 5\epsilon/6 = \epsilon,\end{aligned}$$

where the last inequality is justified by (5.5). Since $t \in \mathbb{T}$ was arbitrary, we conclude that $\|u - \tilde{u}\|_n < \epsilon$ for any $n = 1, \ldots, M$, which concludes the proof. □

Example 5.1.7 (Computability on $C^1(\mathbb{R})$). For the sake of completeness we must also construct a computability structure on the space $\mathcal{Y} = C^1(\mathbb{R})$ of continuously differentiable, real-valued functions. This construction is quite similar to those in Examples 5.1.4 and 5.1.5 and so we shall only sketch it. The countable dense subset $\mathcal{Y}_c$ will consist of continuously differentiable, piecewise quadratic functions. Formally, for each $N \in \mathbb{N}$ and each tuple $(u_0, v_{-N^2}, \ldots, v_{N^2})$, we consider functions $u \in C^1(\mathbb{R}), v \in C(\mathbb{R})$ such that

- $v(x) = v_{-N^2}$ for every $x \leq -N$ and $v(x) = v_{N^2}$ for every $x \geq N$;
- $v(j/N) = p_j$ for each $j = -N^2, \ldots, 0, \ldots, N^2$;
- v is piecewise linear on the interval $[j/N, (j+1)/N]$, for each $j = -N^2, \ldots, 0, \ldots, N^2 - 1$.
- $u(x) = u_0 + \int_0^x v(\xi)d\xi$.

Following the same lines of reasoning as in the previous examples, we can see that the space $\mathcal{Y}_c$ of functions u considered above is a dense countable subset of $\mathcal{Y}$, we can define an enumeration $\alpha_{\mathcal{Y}} : \mathbb{N} \to \mathcal{Y}_c$ and thus, following the construction of Definition 5.1.2, we obtain a computability structure on $\mathcal{Y}_c$.

We mention in passing that v is an element of the countable subset $\mathcal{X}_c$ defined in Example 5.1.4. We also remark that other equivalent, constructible approaches could be used, for example, polynomial interpolation (using as data the values $u_{-N^2}, \dots, u_{N^2}$ of u at collocated points), truncated Taylor series (using as data the values $u_0^{(0)}, \dots, u_0^{(N)}$ of u and its derivatives at zero), or cubic splines (using as data the values $u_{-N^2}, \dots, u_{N^2}, v_{-N^2}, \dots, v_{N^2}$ of u and its first derivative at collocated points).

Example 5.1.8 (**Computability on $\mathcal{X} \times \mathcal{Y}$**)**.** Given computability structures on spaces $\mathcal{X}, \mathcal{Y}$, one can define a computability structure on the product $\mathcal{X} \times \mathcal{Y}$ as follows. If $(\mathcal{X}_c, \alpha_{\mathcal{X}})$ and $(\mathcal{Y}_c, \alpha_{\mathcal{Y}})$ are enumerated countable dense subsets, we consider $(\mathcal{X} \times \mathcal{Y})_c = \mathcal{X}_c \times \mathcal{Y}_c$ as a countable dense subset of $\mathcal{X} \times \mathcal{Y}$ and define the enumeration $\alpha_{\mathcal{X}\times\mathcal{Y}} : \mathbb{N} \to \mathcal{X} \times \mathcal{Y}$ as $\alpha_{\mathcal{X}\times\mathcal{Y}}(\langle \ell, r \rangle) = (\alpha_{\mathcal{X}}(\ell), \alpha_{\mathcal{Y}}(r))$.

We can then apply the construction on Definition 5.1.2 and obtain a set of computable elements $C_{\bar{\alpha}_{\mathcal{X}\times\mathcal{Y}}}$ on $\mathcal{X} \times \mathcal{Y}$. When doing this, we must define a metric on $\mathcal{X} \times \mathcal{Y}$, which can be easily induced[1] from the metrics on $\mathcal{X}$ and $\mathcal{Y}$. The set of computable elements thus constructed can be described in terms of the computability structures on $\mathcal{X}$ and $\mathcal{Y}$; we leave as an exercise to the reader to check that $C_{\bar{\alpha}_{\mathcal{X}\times\mathcal{Y}}} = C_{\bar{\alpha}_{\mathcal{X}}} \times C_{\bar{\alpha}_{\mathcal{Y}}}$.

The procedure described above can be easily generalized to finite products, i.e. to construct a computability structure on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_N$ given computability structures on $\mathcal{X}_1, \dots, \mathcal{X}_N$. In particular, one can use a computability structure on $\mathcal{X}$ to define a computability structure on $\mathcal{X}^N$, for $N \in \mathbb{N}$.

Definition 5.1.9 (**Tracking function**)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complete metric spaces and $(\mathcal{X}_c, \alpha_{\mathcal{X}})$, $(\mathcal{Y}_c, \alpha_{\mathcal{Y}})$ be enumerated countable dense subsets. Let $(\Omega_{\bar{\alpha}_{\mathcal{X}}}, C_{\bar{\alpha}_{\mathcal{X}}}, \bar{\alpha}_{\mathcal{X}})$ and $(\Omega_{\bar{\alpha}_{\mathcal{Y}}}, C_{\bar{\alpha}_{\mathcal{Y}}}, \bar{\alpha}_{\mathcal{Y}})$ be the corresponding computability structures. Let $f : \mathcal{X} \rightharpoonup \mathcal{Y}$ and $\varphi : \mathbb{N} \rightharpoonup \mathbb{N}$.

We say that φ is a *tracking function* with respect to $(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})$, or an $(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})$-*tracking function*, for f, if for all $n \in \Omega_{\bar{\alpha}_{\mathcal{X}}}$,

- if $\bar{\alpha}_{\mathcal{X}}(n) \in \operatorname{dom} f$, then $n \in \operatorname{dom} \varphi$ and $\varphi(n) \in \Omega_{\bar{\alpha}_{\mathcal{Y}}}$ and $f(\bar{\alpha}_{\mathcal{X}}(n)) = \bar{\alpha}_{\mathcal{Y}}(\varphi(n))$;
- if $\bar{\alpha}_{\mathcal{X}}(n) \notin \operatorname{dom} f$, then $n \notin \operatorname{dom} \varphi$.

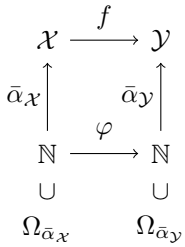


Figure 5.4: Tracking function.

Definition 5.1.10 (**Tracking computability**)**.** Let $\mathcal{X}$ and $\mathcal{Y}$ be complete metric spaces with computability structures as in Definition 5.1.2. We say that a function $f : \mathcal{X} \rightharpoonup \mathcal{Y}$ is *tracking computable* with respect to $(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})$, or $(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})$-*computable*, if it has a computable $(\alpha_{\mathcal{X}}, \alpha_{\mathcal{Y}})$-tracking function $\varphi : \mathbb{N} \rightharpoonup \mathbb{N}$.

[1] For example, we may take the metric to be one of the following: $d_{\mathcal{X}\times\mathcal{Y}}((x_1, y_1), (x_2, y_2)) = d_{\mathcal{X}}(x_1, x_2) + d_{\mathcal{Y}}(y_1, y_2)$; $d_{\mathcal{X}\times\mathcal{Y}}((x_1, y_1), (x_2, y_2)) = \max(d_{\mathcal{X}}(x_1, x_2), d_{\mathcal{Y}}(y_1, y_2))$; $d_{\mathcal{X}\times\mathcal{Y}}((x_1, y_1), (x_2, y_2)) = \sqrt{d_{\mathcal{X}}(x_1, x_2)^2 + d_{\mathcal{Y}}(y_1, y_2)^2}$. These are all equivalent.

5.2 Computability of the $\mathcal{X}$-GPAC modules and induced operators

To illustrate the generality of Definitions 5.1.9 and 5.1.10, we proceed to prove that all the basic modules considered in the $\mathcal{X}$-GPAC (Definition 3.4.2) generate tracking computable functions.

Lemma 5.2.1 (**Tracking computability of the $\mathcal{X}$-GPAC basic modules**). *Let $\mathcal{X} = C(\mathbb{R})$, $\mathcal{Y} = C^1(\mathbb{R})$, $\mathcal{Z} = C^1(\mathbb{T}, \mathcal{X})$ and $\mathcal{W} = C^1(\mathbb{T}, \mathcal{Y})$, with the computability structures described in Examples 5.1.4, 5.1.5 and 5.1.7.*

1. ***constants**: let g be a computable element in $\mathcal{X}$ (that is, $g \in C_{\bar{\alpha}_{\mathcal{X}}}$), then the constant stream $u \in C^1(\mathbb{T}, \mathcal{X})$ given by $u(t) = g$ is a computable element in $C^1(\mathbb{T}, \mathcal{X})$ (that is, $u \in C_{\bar{\alpha}_{\mathcal{Z}}}$);*
2. ***addition**: the following function is $(\alpha_{\mathcal{Z}\times\mathcal{Z}}, \alpha_{\mathcal{Z}})$-computable:*
$$\mathbf{add} : C^1(\mathbb{T}, \mathcal{X}) \times C^1(\mathbb{T}, \mathcal{X}) \to C^1(\mathbb{T}, \mathcal{X}), \text{ given by } \mathbf{add}(u, v)(t) = u(t) + v(t);$$
3. ***multiplication**: the following function is $(\alpha_{\mathcal{Z}\times\mathcal{Z}}, \alpha_{\mathcal{Z}})$-computable:*
$$\mathbf{mult} : C^1(\mathbb{T}, \mathcal{X}) \times C^1(\mathbb{T}, \mathcal{X}) \to C^1(\mathbb{T}, \mathcal{X}), \text{ given by } \mathbf{mult}(u, v)(t) = u(t)v(t);$$
4. ***integration**: the following function is $(\alpha_{\mathcal{X}\times\mathcal{Z}\times\mathcal{Z}}, \alpha_{\mathcal{Z}})$-computable:*
$$\mathbf{int} : \mathcal{X} \times C^1(\mathbb{T}, \mathcal{X}) \times C^1(\mathbb{T}, \mathcal{X}) \to C^1(\mathbb{T}, \mathcal{X}), \text{ given by } \mathbf{int}(g, u, v)(t) = g + \int_0^t u(s)v'(s)ds;$$
5. ***differentiation**: the following function is $(\alpha_{\mathcal{W}}, \alpha_{\mathcal{Z}})$-computable:*
$$\mathbf{diff} : C^1(\mathbb{T}, \mathcal{Y}) \to C^1(\mathbb{T}, \mathcal{X}), \text{ given by } \mathbf{diff}(u)(t) = u'(t).$$

Proof. In this proof, we shall assume for simplicity that $\mathbb{T} = [0, 1]$, but the proof carries out with minor changes for the cases $\mathbb{T} = [0, T]$ and $\mathbb{T} = [0, \infty)$.

Constants: given an element $g \in \mathcal{X}_c$, the constant stream $u(t) = g$ is in $\mathcal{Z}_c$; in particular, it is encoded by $N = 1$ and the tuple $(g, 0, 0)$. Moreover, let $n_0 \in \mathbb{N}$ such that $\alpha_{\mathcal{X}}(n_0) = 0$. Then, given a code $c = \langle e, m\rangle$ for an element $g \in C_{\bar{\alpha}_{\mathcal{X}}}$, consider the code $c' = \langle e', m\rangle$ in which $\{e'\}(n) = \langle 1, \langle\{e\}(n), n_0, n_0\rangle\rangle$; according to Example 5.1.5, this corresponds to $N = 1$, $x_0 = \alpha_{\mathcal{X}}(\{e\}(n))$, $y_0 = \alpha_{\mathcal{X}}(n_0) = 0$ and $y_1 = \alpha_{\mathcal{X}}(n_0) = 0$. Then c' is a code for the desired constant stream, so that $u \in C_{\bar{\alpha}_{\mathcal{Z}}}$.

Refinement: for the next three basic modules, it is useful to prove that a certain operation on the codes of elements in $\mathcal{Z}_c$ (hereby called *refinement*) is computable.

Let $n = \langle N, \langle m_0, m'_0, \ldots, m'_N\rangle\rangle$ be a code for a stream $u \in C^1(\mathbb{T}, \mathcal{X})$. By a *refinement* of n we mean a natural number $\bar{n} = \langle \bar{N}, \langle \bar{m}_0, \bar{m}'_0, \ldots, \bar{m}'_{\bar{N}}\rangle\rangle$ such that $\bar{n}$ is also a code for u and $\bar{N}$ is a multiple of N. Intuitively, this means that we are encoding the stream u with additional (yet redundant) data (see Figure 5.5). We shall see very shortly how the values of $\bar{m}_0, \bar{m}'_j$ must depend in a straightforward way on the values of m_0, m'_j.

Given a code $n = \langle N, \langle m_0, m'_0, \ldots, m'_N\rangle\rangle$ and a positive integer k, we construct a refinement $R(n, k) = \langle \bar{N}, \langle \bar{m}_0, \bar{m}'_0, \ldots, \bar{m}'_N\rangle\rangle$ as follows:

- $\bar{N} = N \times k$;

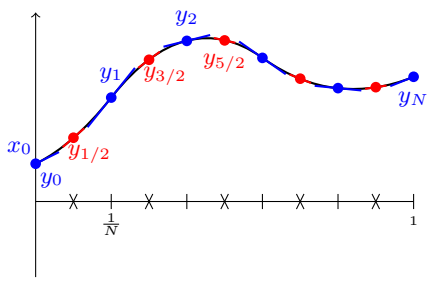


Figure 5.5: A stream $u \in C^1(\mathbb{T}, \mathcal{X})$ with two encodings: one having data $(x_0, y_0, \dots y_N)$ (in blue), and the other being a refinement with roughly twice as much data (the additional data are in red). Furthermore, the refined data can be obtained from the original data; in particular, $y_{1/2} = (y_0 + y_1)/2$, $y_{3/2} = (y_1 + y_2)/2$ and so on.

- $\bar{m}_0 = m_0$;
- for $j = 0, \dots, \bar{N}$, find $0 \leq i < N$ and $0 \leq \ell < k$ such that $j = ki + \ell$; then $\bar{m}'_j$ is chosen such that $\alpha_{\mathcal{X}}(\bar{m}'_j) = \frac{k-\ell}{k}\alpha_{\mathcal{X}}(m'_i) + \frac{\ell}{k}\alpha_{\mathcal{X}}(m'_{i+1})$.

We note that each of these values can be obtained from n in an effective manner: in particular, the construction of $\bar{m}'_j$ is effective since addition and multiplication by rationals is a tracking computable[2] operation in $\mathcal{X} = C(\mathbb{R})$.

Next consider $n_1 = \langle N_1, \langle \dots \rangle\rangle$ and $n_2 = \langle N_2, \langle \dots \rangle\rangle$ two codes for elements in $\mathcal{Z}_c$. We can take the least common multiple of the step numbers, $\bar{N} = \operatorname{lcm}(N_1, N_2)$ and then consider a *common refinement*, that is, a pair of codes $\bar{n}_1 = \langle \bar{N}, \langle \dots \rangle\rangle$, $\bar{n}_2 = \langle \bar{N}, \langle \dots \rangle\rangle$ such that $\bar{n}_1, \bar{n}_2$ are refinements of n_1, n_2. We observe that each step in this construction is effective, and thus the map $(n_1, n_2) \mapsto (\bar{n}_1, \bar{n}_2)$ is computable.

Addition: With the tool of common refinements at our disposal, we can proceed to show that addition is tracking computable. First notice that, if u_1, u_2 are streams in $\mathcal{Z}_c$ given by the same step number $\bar{N}$ and data tuples $(x_{1,0}, y_{1,0}, \dots, y_{1,\bar{N}})$, $(x_{2,0}, y_{2,0}, \dots, y_{2,\bar{N}})$, then their sum $u_1 + u_2$ is a stream in $\mathcal{Z}_c$ given by $\bar{N}$ and the data tuple $(x_{1,0} + x_{2,0}, y_{1,0} + y_{2,0}, \dots, y_{1,\bar{N}} + y_{2,\bar{N}})$. Since addition in $\mathcal{X}$ is tracking computable, this procedure can be effectivized. Formally, given two codes n_1, n_2 for elements $\alpha_{\mathcal{Z}}(n_1), \alpha_{\mathcal{Z}}(n_2) \in \mathcal{Z}_c$, we can effectively obtain a code for its sum $\alpha_{\mathcal{Z}}(n_1) + \alpha_{\mathcal{Z}}(n_2)$ as follows:

- build a common refinement $\bar{n}_1 = \langle \bar{N}, \langle m_{1,0}, m'_{1,0}, \dots m'_{1,\bar{N}} \rangle\rangle$, $\bar{n}_2 = \langle \bar{N}, \langle m_{2,0}, m'_{2,0}, \dots m'_{2,\bar{N}} \rangle\rangle$ of n_1, n_2;
- take the code $n_+ = \langle \bar{N}, \langle m_{+,0}, m'_{+,0}, \dots m'_{+,\bar{N}} \rangle\rangle$, where $\alpha_{\mathcal{X}}(m_{+,0}) = \alpha_{\mathcal{X}}(m_{1,0}) + \alpha_{\mathcal{X}}(m_{2,0})$ and so on.

Let $\mathbf{add}_c : (n_1, n_2) \mapsto n_+$ denote the map that from two codes in $\mathcal{Z}_c$ gives the code for their sum. Now let $c_1 = \langle e_1, m_1 \rangle$, $c_2 = \langle e_2, m_2 \rangle$ be codes for streams $u \in \mathcal{Z}$. In particular, c_1, c_2 encode effective Cauchy sequences in $\mathcal{Z}_c$ converging to computable elements in $C_{\bar{\alpha}_{\mathcal{Z}}}$. We must construct a code $c_+ = \langle e_+, m_+ \rangle$ for their sum:

[2] To see this, the reader may wish to either consult standard texts in computable analysis, such as [Wei00, Chapter 6] or [PER89, Chapter 0], or adapt the current proof to $\mathcal{X} = C(\mathbb{R})$, namely, consider refinements in $\mathcal{X}$ according to the coding of Example 5.1.4.

- take e_+ to be a code for the function $n \mapsto \mathbf{add}_c(\{e_1\}(n), \{e_2\}(n))$, that is, the function that maps each n to a code of the sum $\alpha(\{e_1\}(n))+\alpha(\{e_2\}(n))$ (in other words, the Cauchy sequence encoded by c_+ is the termwise addition of the Cauchy sequences encoded by c_1 and c_2);
- take m_+ to be a code for a function $\{m_+\}$ as follows: for a given $\nu \in \mathbb{N}$, take δ, M, δ', ν' with $\delta \leq 2^{-\nu}/2$; $2^{-M} \leq 2^{-\nu}/2$; $\delta' \leq \delta/2$; and $2^{-\nu'} \leq \delta' 2^{-M}$. Next define $\{m_+\}(\nu) = \max(\{m_1\}(\nu'), \{m_2\}(\nu')) =: N$. Now observe that, for $m_1, m_2 \geq N$,

  $d_{\mathcal{Z}}(\alpha\{e_i\}(m_1), \alpha\{e_i\}(m_2)) < 2^{-\nu'}, i = 1, 2$ (from Definition 4.2.3), so that
  $\|\alpha\{e_i\}(m_1) - \alpha\{e_i\}(m_2)\|_n < \delta', i = 1, 2, n = 1, \dots M$ (from Proposition 2.2.15), so that
  $\|\alpha\{e_+\}(m_1) - \alpha\{e_+\}(m_2)\|_n < \delta, n = 1, \dots M$ (from triangular inequality), so that
  $d_{\mathcal{Z}}(\alpha\{e_+\}(m_1), \alpha\{e_+\}(m_2)) < 2^{-\nu}$ (from Proposition 2.2.15).

We conclude that c_+ is in fact a code for the effective Cauchy sequence corresponding to the addition of the sequences given by c_1 and c_2; in particular, it must converge to the computable element corresponding to the sum of their limits. Moreover, the entire procedure described above can be effectivized, so that we can define a computable function $\phi : (c_1, c_2) \mapsto c_+$. It follows that ϕ is a tracking function for **add**, so that addition in $C^1(\mathbb{T}, \mathcal{X})$ is tracking computable.

Multiplication: there are a lot of technical details that go into defining a tracking function for multiplication. We shall only sketch a proof of this result. In short, there are three 'levels' on which we must *approximately compute* multiplication.

At the first level, we have to approximate multiplication on $\mathcal{X} = C(\mathbb{R})$. Remember that functions in $\mathcal{X}$ are approximated by continuous, piecewise linear, rational functions. These functions are in turn represented by tuples $(x_{-N^2}, \dots, x_{N^2})$ of rational numbers, and encoded as a single natural number.

Therefore, let $g_1 \sim (x^1_{-N^2}, \dots, x^1_{N^2})$ and $g_2 \sim (x^2_{-N^2}, \dots, x^2_{N^2})$ be representations of two functions in $\mathcal{X}_c$, where N specifies a large enough common refinement. We will approximate their product by the function $g_\times \sim (x^\times_{-N^2}, \dots, x^\times_{N^2})$, whose representation is given by $x^\times_j = x^1_j x^2_j$. Observe that, while multiplications in the rationals can be done exactly, multiplication over $\mathcal{X}_c$ is no longer exact; in particular, $g_\times$ is a piecewise linear function whereas the product $g_1 g_2$ is piecewise quadratic. Therefore, there is an approximation error, which with some effort can be found to be bounded as

$$\|g_\times - g_1 g_2\|_N \leq C \sup |x^1_{j+1} - x^1_j| \sup |x^2_{j+1} - x^2_j|, \tag{5.6}$$

for the constant $C = 1/4$. It is worth noticing that $|x^i_{j+1} - x^i_j|$ correspond to differences in consecutive points of the representations of g_i. This difference can in principle be effectively bounded by the value of the discretization size N (but we omit these details).

At the second level, we can now consider multiplication in $\mathcal{Z}_c$, where $\mathcal{Z} = C^1(\mathbb{T}, \mathcal{X})$. Let $u_1 \sim (f^1_0, g^1_0, \dots, g^1_N)$ and $u_2 \sim (f^2_0, g^2_0, \dots, g^2_N)$ be two representations of functions in $\mathcal{Z}_c$ (again, we take a large enough common refinement value N). Note our notation: we are denoting $f^i_0 := u_i(0)$ and $g^i_j := u_i'(j/N)$. To approximate their product we recall the product rule for derivatives, $(u_1 u_2)'(t) = u_1(t)u_2'(t) + u_1'(t)u_2(t)$. To compute this expression at collocated values of t, we must first find the values of $f^i_j := u_i(j/N)$. Since u_i are piecewise quadratic, these values can be recursively obtained by integration using the trapezoid rule,

$$f^i_{j+1} = f^i_j + \frac{\Delta t}{2}(g^i_j + g^i_{j+1}), \quad \Delta t = \frac{1}{N}. \tag{5.7}$$

We observe that the values of f^i_j can be *exactly* computed over $\mathcal{X}_c$ from the representation of u_i.

Therefore, we can approximate the product $u_1 u_2$ by the function $u_\times \sim (f_0^\times, g_0^\times, \dots, g_N^\times)$, where

- $f_0^\times$ is the approximate computation (over $\mathcal{X}_c$) of $f_0^1 f_0^2$;
- $g_j^\times$ is the approximate computation (over $\mathcal{X}_c$) of $f_j^1 g_j^2 + f_j^2 g_j^1$.

Observe that these computations are performed at the first level, and have corresponding approximation errors as in (5.6). Once again, multiplication over $\mathcal{Z}_c$ is not exact; in particular, $u_\times$ is piecewise quadratic whereas $u_1 u_2$ is piecewise quartic. It takes considerably more effort to describe an effective bound on the approximation error, which will depend on: the approximation error for the first level; the time step size $\Delta t = 1/N$; the norms of $\|u_i\|_m \approx \sup \|f_j^i\|_m + \sup \|g_j^i\|_m$; and the consecutive differences $\sup \|f_{j+1}^i - f_j^i\|_m$, $\sup \|g_{j+1}^i - g_j^i\|_m$. Ultimately, we can bound this error in an effective way depending on the refinement value N.

Finally, at the third level we can consider multiplication in $\mathcal{Z}$. Let $c_1 = \langle e_1, m_1 \rangle$ and $c_2 = \langle e_2, m_2 \rangle$ be codes for effective Cauchy sequences in $\mathcal{Z}_c$ converging to a computable element in $\mathcal{Z}$. We wish to define a code $c_\times = \langle e_\times, m_\times \rangle$ for an effective Cauchy sequence converging to their product. We can take $e_\times$ to encode the termwise product of the two sequences, so that $\alpha_{\mathcal{Z}}(\{e_\times\}(n))$ is an approximation of $\alpha_{\mathcal{Z}}(\{e_1\}(n))\alpha_{\mathcal{Z}}(\{e_2\}(n))$, computed at the second level (described above). We must also require that later terms in the sequence are computed with higher accuracy. For example, we require that $\{e_\times\}(n)$ is computed using a common refinement of size at least n. The specification of the modulus of convergence $\{m_\times\}$ takes more effort. Given ν, we want to find a value for $\{m_\times\}(\nu)$ according to Definition 4.2.3. We sketch the desired procedure:

- let $u_{1,n} = \alpha_{\mathcal{Z}}(\{e_1\}(n))$, $u_{2,n} = \alpha_{\mathcal{Z}}(\{e_2\}(n)$ and $u_{\times,n} = \alpha_{\mathcal{Z}}(\{e_\times\}(n))$;
- find an uniform bound (i.e. independent of n) on $\|u_{1,n}\|_m$ and $\|u_{2,n}\|_m$ (possible since these are effective Cauchy sequences);
- take into account the inequality

$$\|u_{1,n}u_{2,n} - u_{1,n'}u_{2,n'}\|_m \leq \|u_{1,n}\|_m \|u_{2,n} - u_{2,n'}\|_m + \|u_{2,n'}\|_m \|u_{1,n} - u_{1,n'}\|_m;$$

  use this to get an effective bound on $\|u_{1,n}u_{2,n} - u_{1,n'}u_{2,n'}\|_m$;
- take into account the inequality

$$\|u_{\times,n} - u_{\times,n'}\|_m \leq \|u_{\times,n} - u_{1,n}u_{2,n}\|_m + \|u_{1,n}u_{2,n} - u_{1,n'}u_{2,n'}\|_m + \|u_{\times,n'} - u_{1,n'}u_{2,n'}\|_m;$$

  two of the terms on the right-hand side correspond to the approximation errors from the second level, whereas the third term can be bounded from the previous step;
- find a large enough value of N such that all three terms in the previous step are small enough for $n, n' \geq N$;
- repeat the above steps for pseudonorm indices $m = 1, \dots, M$, where Proposition 2.2.15 is taken into account, to retrieve the desired bound on $d(u_{\times,n}, u_{\times,n'})$. Then $\{m_\times\}(\nu)$ can be taken to be the largest value of N required.

Integration: The techniques and ideas used for defining a tracking function for multiplication can be easily adapted into integration as well. Suppose that $g \in \mathcal{X}_c$ and $u, v \in \mathcal{Z}_C$. Moreover let u, v be represented by the tuples of data $(f_0^1, g_0^1, \dots, g_N^1)$ and $(f_0^2, g_0^2, \dots, g_N^2)$, respectively, where a large enough common refinement N is taken. We recall the notation $f_j^1 = u(j/N)$ for the values

of u at collocated points, which can be recursively and *exactly* computed by (5.7). Now let $w \in \mathcal{Z}$ be given by $w(t) = g + \int_0^t u(s)dv(s)$. Observe that $w(0) = g$ and $w'(t) = u(t)v'(t)$; in particular, $w'(j/N) = f_j^1 g_j^2$. Therefore, we can approximate w by the function $w_\int \sim (f_0^\int, g_0^\int, \ldots, g_N^\int)$ in $\mathcal{Z}_c$, where

- $f_0^\int = g$;
- $g_j^\int$ is the approximate computation (over $\mathcal{X}_c$) of $f_j^1 g_j^2$.

Similarly to the previous case, these computations have approximation errors as in (5.6). We also need to estimate the approximation error between w and $w_\int$ in an effective way depending on the refinement parameter N. Finally, we extend the computation for codes $c_0 = \langle e_0, m_0\rangle$, $c_1 = \langle e_1, m_1\rangle$, $c_2 = \langle e_2, m_2\rangle$ of effective Cauchy sequences converging to g, u, v respectively. We define a code $c_\int = \langle e_\int, m_\int\rangle$ for an effective Cauchy sequence converging to the integral. We take $e_\int$ to be the code of a recursive function such that

$$\alpha_{\mathcal{Z}}(\{e_\int\}(n)) \text{ is the approximation of } \mathbf{int}(\alpha_{\mathcal{X}}(\{e_0\}(n)), \alpha_{\mathcal{Z}}(\{e_1\}(n)), \alpha_{\mathcal{Z}}(\{e_2\}(n))),$$

computed in a common refinement of size at least n. The specification of the modulus of convergence is done in a similar way as for multiplication and we omit the details.

Differentiation: The case of differentiation is actually much simpler to treat. First observe that, if g is a function in $\mathcal{Y}_c$ represented by $(x_0, y_{-N^2}, \ldots, y_{N^2})$, then its derivative g' is a function in $\mathcal{X}_c$ represented by $(y_{-N^2}, \ldots, y_{N^2})$. Thus, differentiation in $\mathcal{Y}_c$ can be *exactly* computed. Moreover, if u is a function in $\mathcal{W}_c$ represented by $(f_0, g_0, \ldots, g_N)$, then its spatial derivative is a function in $\mathcal{Z}_c$ represented by $(f_0', g_0', \ldots, g_N')$ (note that each $f_0, g_j \in \mathcal{Y}$). Thus, differentiation in $\mathcal{W}_c$ can be *exactly* computed as well. Finally, if $c = \langle e, m\rangle$ is a code for an effective Cauchy sequence converging to a computable element in $\mathcal{W}$, we can easily define a code $c_\delta = \langle e_\delta, m_\delta\rangle$ for an effective Cauchy sequence converging to its spatial derivative. Just take e_δ such that $\alpha_{\mathcal{Z}}(\{e_\delta\}(n))$ is the (exact) derivative of $\alpha_{\mathcal{W}}(\{e\}(n))$ and we can even use the same modulus of convergence, $m_\delta = m$. This is because, if $g \in \mathcal{W}$, then for each n

$$\|g'\|_{\mathcal{Y},n} = \sup \|g'(t)\|_{\mathcal{X},n} \leq \sup \|g(t)\|_{\mathcal{X},n} + \sup \|g'(t)\|_{\mathcal{X},n} = \|g\|_{\mathcal{W},n}$$

and thus $d_{\mathcal{Y}}(u', v') \leq d_W(u, v)$ for any $u, v \in \mathcal{W}$. □

Remark 5.2.2. It is important to note that our treatment of the differential module is somewhat different from that presented on Chapter 3. We have proven that the *total-valued* differential functional of type $C^1(\mathbb{T}, C^1(\mathbb{R})) \to C^1(\mathbb{T}, C(\mathbb{R}))$ is tracking computable. We did not wish to study tracking computability of the *partial-valued* differential functional of type $C^1(\mathbb{T}, C(\mathbb{R})) \rightharpoonup C^1(\mathbb{T}, C(\mathbb{R}))$, since this operator is not continuous. As a consequence, the definitions of induced operator and generable functions will be slightly different in this chapter. These will be given shortly at the start of Section 5.3.

Next we consider the continuous limit module studied in Chapter 4 and show its tracking computability. Note that the limit operation only makes sense when $\mathbb{T} = [0, \infty)$.

Lemma 5.2.3 (Tracking computability of the continuous limit module). *Let $\mathcal{X} = C(\mathbb{R})$, $\mathbb{T} = [0, \infty)$ and $\mathcal{Z} = C^1(\mathbb{T}, \mathcal{X})$, with the computability structures described in Examples 5.1.4 and 5.1.5. Consider the function*

$$\mathbf{lim} : C^1(\mathbb{T}, \mathcal{X}) \rightharpoonup \mathcal{X}, \text{ given by } \mathbf{lim}(u) = \lim_{t\to\infty} u(t),$$

whose domain consists of id*-convergent Cauchy streams* $u \in C^1(\mathbb{T}, \mathcal{X})$ *(cf. Definitions 4.2.3 and 4.2.4). Then* **lim** *is* $(\alpha_{\mathcal{Z}}, \alpha_{\mathcal{X}})$*-computable.*

Proof. If $u_n \in C^1(\mathbb{T}, \mathcal{X})$ is an effective Cauchy sequence converging to a stream $u \in C^1(\mathbb{T}, \mathcal{X})$ which in turn has an id-convergent limit $f \in \mathcal{X}$, then $\lim_{t\to\infty} \lim_{n\to\infty} u_n(t) = f$. Thus, a candidate for an approximation of f would be given by $u_{k_n}(t_n)$, where t_n and k_n are large enough natural numbers. Our goal is to somehow effectivize this line of thought.

Let $u \in \mathcal{Z}_c$. Since $\mathbb{T} = [0, \infty)$, a representation of u must be given by $(f_0, g_0, \dots g_{N^2})$, where each $f_0, g_j \in \mathcal{X}_c$. Recall that this notation means that $f_0 = u(0)$ and $g_j = u'(j/N)$ for $j = 0, \dots, N^2$. Our first observation is that, for any natural number $n \in \mathbb{N}$, the value of $u(n)$ can be *exactly* computed from a representation of u: in particular,

$$u(n) = \begin{cases} f_{nN} & \text{if } n \leq N; \\ f_{N^2} + (N-n)g_{N^2} & \text{if } n \geq N, \end{cases}$$

where the functions f_j can be exactly computed as in (5.7) (but with $\Delta t = 1/N^2$).

Now let $c = \langle e, N\rangle$ be a code for an effective Cauchy sequence in $\mathcal{Z}_c$ converging to a computable element $u \in C^1(\mathbb{T}, \mathcal{X})$. We want to compute a code $c_\ell = \langle e_\ell, N_\ell\rangle$ for an effective Cauchy sequence in $\mathcal{X}_c$ converging to the limit $f = \mathbf{lim}(u) \in \mathcal{X}$. Let $u_n = \alpha_{\mathcal{Z}}(\{e\}(n))$ and $f_n = \alpha_{\mathcal{X}}(\{e_\ell\}(n))$. As mentioned before, we shall construct e_ℓ in such a way that

$$f_n = u_{k_n}(t_n), \text{ for suitable choices of } k_n, t_n;$$

later on, it will be revealed that $t_n = n + 1$ and $k_n = \{N\}(2n + 7)$ are the suitable choices.

We desire f_n to be an effective Cauchy sequence. Thus let $\nu \in \mathbb{N}$ be given, and we want to find K such that, for $n, m \geq K$ one has $d_{\mathcal{X}}(f_n, f_m) < 2^{-\nu}$. By applying the triangular inequality we can write

$$d_{\mathcal{X}}(f_n, f_m) \leq d_{\mathcal{X}}(u_{k_n}(t_n), u(t_n)) + d_{\mathcal{X}}(u(t_n), u(t_m)) + d_{\mathcal{X}}(u(t_m), u_{k_m}(t_m)).$$

Since u is an id-convergent Cauchy stream, we can bound the second term of the above sum. Namely, take $\tau = \nu + 1$, and thus if $t_n, t_m \geq \tau$, then $d_{\mathcal{X}}(u(t_n), u(t_m)) < 2^{-\tau} = 2^{-\nu-1}$. Next we need to handle the term $d_{\mathcal{X}}(u_{k_n}(t_n), u(t_n))$, which amounts to finding a suitably large k_n.

We shall apply Proposition 2.2.15 to convert between metric and pseudonorms. Let $\epsilon, \delta, M, \delta', \nu'$ be constructed as follows: $\epsilon = 2^{-n}/4 = 2^{-n-2}$, $\delta = \epsilon/2 = 2^{-n-3}$, $2^{-M} = \epsilon/2$ (so that $M = n + 3$), $\delta' = \delta 2^{-M} = 2^{-2n-6}$ and $2^{-\nu'} = \delta'/2$ (so that $\nu' = 2n + 7$). Finally, take $k_n = \{N\}(\nu')$. If $m \geq k_n$ we have that $d_{\mathcal{Z}}(u_{k_n}, u_m) < 2^{-\nu'}$. Since $u_m \to u$ this implies that $d_{\mathcal{Z}}(u_{k_n}, u) \leq 2^{-\nu'} < \delta'$. Applying Proposition 2.2.15 we conclude that $\|u_{k_n} - u\|_{\mathcal{Z},M} < \delta$. Let $t_n = n+1$ and observe that $0 \leq t_n \leq M$. Therefore,

$$\begin{aligned} \|u_{k_n}(t_n) - u(t_n)\|_{\mathcal{X},M} &\leq \sup_{0\leq t\leq M} \|u_{k_n}(t) - u(t)\|_{\mathcal{X},M} \\ &\leq \sup_{0\leq t\leq M} \|u_{k_n}(t) - u(t)\|_{\mathcal{X},M} + \sup_{0\leq t\leq M} \|u'_{k_n}(t) - u'(t)\|_{\mathcal{X},M} \\ &= \|u_{k_n} - u\|_{\mathcal{Z},M} < \delta. \end{aligned}$$

Applying Proposition 2.2.15 once more, we conclude that $d_{\mathcal{X}}(u_{k_n}(t_n), u(t_n)) < \epsilon = 2^{-n-2}$. This reasoning also proves that $d_{\mathcal{X}}(u_{k_m}(t_m), u(t_m)) < 2^{-m-2}$.

To conclude the bound on $d_{\mathcal{X}}(f_n, f_m)$, let $\nu \in \mathbb{N}$ and $n, m \geq \nu$. Since $t_n = n + 1, t_m = m + 1$, it follows that $t_n, t_m \geq \nu + 1$, and thus $d_{\mathcal{X}}(u(t_n), u(t_m)) < 2^{-\nu-1}$. Also, from our previous reasoning, it follows that both $d_{\mathcal{X}}(u_{k_n}(t_n), u(t_n))$ and $d_{\mathcal{X}}(u_{k_m}(t_m), u(t_m))$ are less than $2^{-\nu-2}$. Thus, $d_{\mathcal{X}}(f_n, f_m) < 2^{-\nu-2} + 2^{-\nu-1} + 2^{-\nu-2} = 2^{-\nu}$, as desired. Therefore, f_n is an effective Cauchy

sequence with modulus of convergence equal to the identity (so that we may take N_ℓ to be a code for $\{N_\ell\}(n) = n$).

The construction of t_n and k_n is effective on n and, as mentioned before, the evaluation of a function $u \in \mathcal{Z}_c$ at a natural number $n \in \mathbb{T}$ is also effective (on n and a code for u). Therefore, $\alpha_{\mathcal{X}}(\{e_\ell\}(n)) = f_n = u_{k_n}(t_n)$ is a computable sequence and $c = \langle e, N\rangle \mapsto c_\ell = \langle e_\ell, N_\ell\rangle$ is an effective procedure. Finally, we have proved that, for all $n \in \mathbb{N}$, $d_{\mathcal{X}}(f_n, u(t_n)) < 2^{-n-2}$; since $u(t) \to f$ it also follows that $f_n \to f$, so that c_ℓ is indeed a code for an effective Cauchy sequence converging to $\mathbf{lim}(u)$, as desired. □

Remark 5.2.4. Recall that, in Chapter 4, we actually defined two variants of the continuous limit module: the one-input limit module (for id-convergent Cauchy streams) and the two-input limit module (where the rate of convergence is provided as an input). Clearly, we expect both versions to be equivalent (in fact, we have shown how to derive the one-input limit module from the two-input version), and thus it should be possible to prove a variant of Lemma 5.2.3 for the two-input limit module. However, we won't provide a direct proof of this result. Instead, the tracking computability of the two-input limit module will follow from the fact that many algebraic operations preserve computability, as we shall now see.

Lemma 5.2.5 (**Tracking computability of some algebraic operations**). *The following constructs preserve tracking computability.*

1. ***serial composition**: if $f : \mathcal{X} \rightharpoonup \mathcal{Y}$ and $g : \mathcal{Y} \rightharpoonup \mathcal{Z}$ are tracking computable, so is $g \circ f : \mathcal{X} \rightharpoonup \mathcal{Z}$, defined by $(g \circ f)(x) = g(f(x))$;*
2. ***parallel composition**: if $f : \mathcal{X}_1 \rightharpoonup \mathcal{Y}_1$ and $g : \mathcal{X}_2 \rightharpoonup \mathcal{Y}_2$ are tracking computable, so is $f \times g : \mathcal{X}_1 \times \mathcal{X}_2 \rightharpoonup \mathcal{Y}_1 \times \mathcal{Y}_2$, defined by $(f \times g)(x_1, x_2) = (f(x_1), g(x_2))$;*
3. ***projection**: if $f : \mathcal{X} \rightharpoonup \mathcal{Y}_1 \times \mathcal{Y}_2$ is tracking computable, so are $f_1 : \mathcal{X} \rightharpoonup \mathcal{Y}_1$ and $f_2 : \mathcal{X} \rightharpoonup \mathcal{Y}_2$, defined by $(f_1(x), f_2(x)) = f(x)$.*
4. ***coupling**: if $f_1 : \mathcal{X} \rightharpoonup \mathcal{Y}_1$ and $f_2 : \mathcal{X} \rightharpoonup \mathcal{Y}_2$ are tracking computable, so is $f : \mathcal{X} \rightharpoonup \mathcal{Y}_1 \times \mathcal{Y}_2$, defined by $f(x) = (f_1(x), f_2(x))$.*

Proof. The computability of basic algebraic operations is usually one of the first results to be proven for a model of computation. For example, in the framework of computable analysis, this is proven in [PER89, Section 0.4]; and in the framework of type-2 theory of effectivity, this is proven in [Wei00, Section 2.1].

The proofs are straightforward; we will present the proof of computability for serial composition as an illustration. Let $f : \mathcal{X} \rightharpoonup \mathcal{Y}$ and $g : \mathcal{Y} \rightharpoonup \mathcal{Z}$ be tracking computable, and let $F, G : \mathbb{N} \rightharpoonup \mathbb{N}$ be their (computable) tracking functions. Now $G \circ F$ is a computable function on the natural numbers; we shall prove that $G \circ F$ is a tracking function for $g \circ f$ (using Definition 5.1.9). Let $n \in \Omega_{\bar{\alpha}_{\mathcal{X}}}$ and $x = \bar{\alpha}_{\mathcal{X}}(n)$; we must consider two possibilities.

- Suppose that $x \in \mathrm{dom}(g \circ f)$; then $x \in \mathrm{dom}\, f$, so we can define $y = f(x)$; and also $y \in \mathrm{dom}\, g$. By tracking computability of f it follows that $n \in \mathrm{dom}\, F$ and $y = \bar{\alpha}_{\mathcal{Y}}(F(n))$. By tracking computability of g it follows that $F(n) \in \mathrm{dom}\, G$ and $g(y) = \bar{\alpha}_{\mathcal{Z}}(G(F(n)))$. Thus $n \in \mathrm{dom}(G \circ F)$ and $g \circ f(\bar{\alpha}_{\mathcal{X}}(n)) = \bar{\alpha}_{\mathcal{Z}}(G \circ F(n))$.
- Suppose that $x \notin \mathrm{dom}(g \circ f)$; then either $x \notin \mathrm{dom}\, f$ or $y = f(x)$ and $y \notin \mathrm{dom}\, g$. In the first case, by tracking computability of f it follows that $x \notin \mathrm{dom}\, F$. In the second case, tracking computability of f implies that $n \in \mathrm{dom}\, F$ and $y = \bar{\alpha}_{\mathcal{Y}}(F(n))$; and tracking computability of g implies that $F(n) \notin \mathrm{dom}\, G$. In either way we have that $n \notin \mathrm{dom}(G \circ F)$.

Thus $G \circ F$ is indeed a tracking function for $g \circ f$. We conclude that the serial composition is tracking computable. □

Finally, we can piece all our results together and prove the tracking computability of the induced operator of an LGPAC.

Theorem 22 (**Tracking computability of GPAC induced operators**). *Let $\mathcal{G}$ be an LGPAC (over $\mathcal{X} = C(\mathbb{R})$) with induced operator $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \tilde{\mathcal{M}} \times \mathcal{O}$. Then Φ is tracking computable.*

Proof. We begin by writing

$$\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\tilde{\mathbf{u}}^M, \mathbf{u}^O) = (\Phi_1(\mathbf{inp}_1), \ldots, \Phi_N(\mathbf{inp}_N)); \tag{5.8}$$

in this notation, N is the number of modules in $\mathcal{G}$ and Φ_i are the corresponding functions. For example, if the i-th module is an integrator, then $\mathbf{inp}_i$ is of the form $(g, u, v) \in \mathcal{X} \times C^1(\mathbb{T}, \mathcal{X}) \times C^1(\mathbb{T}, \mathcal{X})$, where g, u, v are the corresponding components of $\mathbf{g}$, $\mathbf{u}^I$ or $\mathbf{u}^M$, and $\Phi_i = \mathbf{int}$; similarly for the other cases.

There is a technical observation we must make at this point. The induced operator is defined slightly differently from Chapter 3, due to the treatment of differential modules (cf. Remark 5.2.2). So if the i-th module is a differential, then $\mathbf{inp}_i$ is a component of $\mathbf{u}^I$ or $\mathbf{u}^M$ belonging to the space $C^1(\mathbb{T}, C^1(\mathbb{R}))$. In other words, the input / mixed spaces may be of the form $C^1(\mathbb{T}, C(\mathbb{R}))^{q_1} \times C^1(\mathbb{T}, C^1(\mathbb{R}))^{q_2}$, instead of the form $C^1(\mathbb{T}, C(\mathbb{R}))^q$ used in Chapter 3 (see in particular Definition 3.4.5). This is the reason for requiring a new symbol $\tilde{\mathcal{M}}$ appearing in the description of the induced operator. We can say that a mixed channel may have a different semantic if it is treated as an input instead of an output. Owing to the inclusion of $C^1(\mathbb{R})$ in $C(\mathbb{R})$, we can say that $\mathcal{M} \subseteq \tilde{\mathcal{M}}$; moreover, the inclusion map $\iota : \mathcal{M} \subseteq \tilde{\mathcal{M}}$ is continuous and tracking computable.

In any case, we now see that each Φ_i described in (5.8) is one of the functions considered in Lemmas 5.2.1 and 5.2.3, and is thus tracking computable. Since Φ can be obtained as the parallel composition and coupling of the Φ_i, we can use Lemma 5.2.5 to conclude that it must also be tracking computable. □

5.3 Tracking computability of LGPAC-generable functions

In this section we prove the main result of this chapter. The goal is to find out under which conditions the function generated by an LGPAC is tracking computable. We recall that, in our terminology, an LGPAC induces an operator

$$\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \tilde{\mathcal{M}} \times \mathcal{O}, \quad \Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\tilde{\mathbf{u}}^M, \mathbf{u}^O);$$

for the LGPAC to generate a valid function, we require the fixed point problem to be well-posed, that is, a map $F : (\mathbf{g}, \mathbf{u}^I) \mapsto (\mathbf{u}^M, \mathbf{u}^O)$ exists, is unique and is continuous.

Previously on Chapter 3, we only required F to be a closed operator (not necessarily continuous). We mentioned that this condition could be strengthened by considering a different topology on the input space given by the graph norm. In fact, for this chapter we shall do exactly that. In other words, to each input channel of a *differential module* we associate the space $C^1(\mathbb{T}, C^1(\mathbb{R}))$ of streams in $C^1(\mathbb{R})$. Under this topology, it turns out that the induced operator becomes continuous. Therefore, we shall require well-posedness for the generable functions. As the resulting Definition is sligthly different than that of Definition 3.4.7, we include it here for clarity.

Definition 5.3.1 (Well-posedness and semantics of LGPAC). Let $\mathcal{G}$ be an LGPAC and $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \tilde{\mathcal{M}} \times \mathcal{O}$ be its induced operator. Let U be an open subset of $\mathcal{C} \times \mathcal{I}$. We say that $\mathcal{G}$ is *well-posed* on U if

- (*existence*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, there exists $(\mathbf{u}^M, \mathbf{u}^O) \in \mathcal{M} \times \mathcal{O}$ such that

$$\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M) = (\mathbf{u}^M, \mathbf{u}^O); \tag{5.9}$$

- (*uniqueness*) for every $(\mathbf{g}, \mathbf{u}^I) \in U$, the tuple $(\mathbf{u}^M, \mathbf{u}^O)$ such that (5.9) holds is unique;
- (*continuity*) the map $F : (\mathbf{g}, \mathbf{u}^I) \mapsto (\mathbf{u}^M, \mathbf{u}^O)$, with domain U and codomain $\mathcal{M} \times \mathcal{O}$, given as the unique solution of (5.9), is continuous.

Under the above conditions, we say that F is the specification of $\mathcal{G}$, or that $\mathcal{G}$ generates F, or that F is LGPAC-*generable*.

Our goal is to find conditions on $\mathcal{G}$ that imply that F is tracking computable. The idea is to find F by solving an *approximate fixed point problem*

$$\text{Given } (\mathbf{g}, \mathbf{u}^I) \text{ and } \epsilon > 0, \text{ find } (\mathbf{u}^M, \mathbf{u}^O) \text{ such that } d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) < \epsilon.$$

Moreover, from the point of view of tracking computability, we look for desired $\mathbf{u}^M, \mathbf{u}^O$ in the enumerated, countable dense subset. Then, by using a sequence of ϵ converging to 0, and under additional assumptions on F (namely, we will require some notion of *effective well-posedness*), this yields a sequence of $\mathbf{u}^M, \mathbf{u}^O$ converging to the desired $F(\mathbf{u}^M, \mathbf{u}^O)$.

Let us now focus on the first step of this construction. Namely, we prove that it is possible to construct *approximate fixed points*.

Lemma 5.3.2. *Let $\mathcal{G}$ be an LGPAC with induced operator $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \to \tilde{\mathcal{M}} \times \mathcal{O}$. Let U be an open subset of $\mathcal{C} \times \mathcal{I}$ and suppose that $\mathcal{G}$ is well-posed on U. Then there exists a procedure $(n, \ell) \mapsto m$ such that, if n is the code for an element in U, that is, $\bar{\alpha}_{\mathcal{C}\times\mathcal{I}}(n) = (\mathbf{g}, \mathbf{u}^I) \in U$ and $\ell \in \mathbb{N}$, then m is the code for an enumerated element in $\mathcal{M} \times \mathcal{O}$, that is, $\alpha_{\mathcal{M}\times\mathcal{O}}(m) = (\mathbf{u}^M, \mathbf{u}^O) \in \mathcal{M} \times \mathcal{O}$; and also $d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) < 2^{-\ell}$.*

Proof. The procedure works as follows. For a given input n, ℓ, let us write $(\mathbf{g}, \mathbf{u}^I) = \bar{\alpha}_{\mathcal{C}\times\mathcal{I}}(n)$. We perform the following dovetailing loop for $m \in \mathbb{N}$:

1. Find m_1, m_2 such that $\alpha_{\mathcal{M}\times\mathcal{O}}(m) = (\alpha_{\mathcal{M}}(m_1), \alpha_{\mathcal{O}}(m_2))$. By our construction (see Example 5.1.8), these can be obtained via the pairing bijections. Also, for clarity, let us write $\mathbf{u}^M = \alpha_{\mathcal{M}}(m_1)$, $\mathbf{u}^O = \alpha_{\mathcal{O}}(m_2)$.
2. Find n' such that $\bar{\alpha}_{\mathcal{C}\times\mathcal{I}\times\mathcal{M}}(n') = (\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M)$. Observe that $\mathbf{u}^M$ is an element in the enumerated, countable dense subset of $\mathcal{M}$, but it can be encoded (as a computable element) by some $\langle e_1, M_1 \rangle$, where e_1 is a code for the constant function $\{e_1\}(n) = m_1$ and M_1 is a code for the identity function. Then, using the pairing functions, the desired n' can be obtained.
3. Find $m' = \varphi(n') = \langle e', M' \rangle$, where φ is a tracking function for Φ. Here we use the fact that Φ is tracking computable (Theorem 22). Notice that

$$\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m') = \bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(\varphi(n')) = \Phi(\bar{\alpha}_{\mathcal{C}\times\mathcal{I}\times\mathcal{M}}(n')) = \Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M).$$

4. Find $m'' = \{e'\}(\{M'\}(\ell+2))$. Observe that, because $\{M'\}$ is a module of convergence for the (Cauchy) sequence $\left(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(\{e'\}(n))\right)_n$, it follows that for $k \geq \{M'\}(\ell+2)$, one has $d(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m''), \alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(\{e'\}(k))) < 2^{-\ell-2}$. In particular, since $\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m')$ is the limit of $\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(\{e'\}(n))$, then $d(\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'), \alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'')) \leq 2^{-\ell-2}$. For clarity, let us write $(\tilde{\mathbf{u}}^M, \tilde{\mathbf{u}}^O) = \alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'')$.

5. Check if $d(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m''), \alpha_{\mathcal{M}\times\mathcal{O}}(m)) < 2^{-\ell-1}$; if yes, then break the loop and return m. Observe that the distance function is tracking computable (to get a close enough approximation, a truncated sum on (5.1) is enough) and it must be evaluated on $\tilde{\mathcal{M}}\times\mathcal{O}$, which is feasible since the inclusion $\mathcal{M} \hookrightarrow \tilde{\mathcal{M}}$ is tracking computable.

Observe that, for some values of m, the corresponding execution of the loop may not terminate. This may happen if $\bar{\alpha}_{\mathcal{C}\times\mathcal{I}\times\mathcal{M}}(n')$ (constructed on step 2) is not an element in the domain of Φ, so that $\varphi(n')$ is a divergent computation, or if $d(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m''), \alpha_{\mathcal{M}\times\mathcal{O}}(m))$ is exactly $2^{-\ell-1}$ (equality may not be a computable predicate). However, if a certain value of m happens to pass our test, then that value satisfies the desired property: indeed,

$$\begin{aligned} d(\Phi(g, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) &\leq d(\Phi(g, \mathbf{u}^I, \mathbf{u}^M), (\tilde{\mathbf{u}}^M, \tilde{\mathbf{u}}^O)) + d((\tilde{\mathbf{u}}^M, \tilde{\mathbf{u}}^O), (\mathbf{u}^M, \mathbf{u}^O)) \\ &= d(\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'), \alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'')) + d(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m''), \alpha_{\mathcal{M}\times\mathcal{O}}(m)) \\ &< 2^{-\ell-2} + 2^{-\ell-1} = 2^{-\ell}. \end{aligned}$$

Moreover, such a value of m can always be found by our algorithm, and to prove this we use the well-posedness of $\mathcal{G}$. Given n such that $(\mathbf{g}, \mathbf{u}^I) = \bar{\alpha}_{\mathcal{C}\times\mathcal{I}}(n) \in U$, we know by well-posedness of $\mathcal{G}$ that there exists (a unique) $(\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}}) \in \mathcal{M}\times\mathcal{O}$ with $\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M_{\mathcal{X}}) = (\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}})$, and thus $d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M_{\mathcal{X}}), (\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}})) = 0$. Now the left hand side of this equality is a continuous expression in $\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}}$ (since Φ and d are continuous), and thus there exists $\delta > 0$ such that for any $\mathbf{u}^M, \mathbf{u}^O \in \mathcal{M}\times\mathcal{O}$ one has

$$\text{if } d((\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}}), (\mathbf{u}^M, \mathbf{u}^O)) < \delta \text{ then } d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) < 2^{-\ell-2}.$$

By density of the enumerated subset, there exists $m \in \mathbb{N}$ such that $\alpha_{\mathcal{M}\times\mathcal{O}}(m) = (\mathbf{u}^M, \mathbf{u}^O)$ with $d((\mathbf{u}^M_{\mathcal{X}}, \mathbf{u}^O_{\mathcal{X}}), (\mathbf{u}^M, \mathbf{u}^O)) < \delta$, and thus $d(F(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) < 2^{-\ell-2}$, or in other words, $d(\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'), \alpha_{\mathcal{M}\times\mathcal{O}}(m)) < 2^{-\ell-2}$. Moreover, the value of m'', computed on step 4, will be such that $d(\bar{\alpha}_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'), \alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m'')) \leq 2^{-\ell-2}$, and a simple application of the triangle inequality yields that $d(\alpha_{\tilde{\mathcal{M}}\times\mathcal{O}}(m''), \alpha_{\mathcal{M}\times\mathcal{O}}(m)) < 2^{-\ell-1}$, so the condition on step 5 is met. Thus the dovetailing loop will effectively succeed in finding a valid m. □

We have shown that it is possible to find approximate fixed points of the induced operator. In the next step, we try to show that *approximate fixed points* are in fact 'close' to *exact fixed points*, which is by no means a trivial statement (rather, there is extensive research on this problem; see for example [KL03, KL10]). Intuitively, we want to establish conditions on the induced operator Φ (and the corresponding solution functional F) that effectively ensure the following: for each ϵ there is δ such that

$$\text{if } d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)) < \delta, \text{ then } d(F(\mathbf{g}, \mathbf{u}^I), (\mathbf{u}^M, \mathbf{u}^O)) < \epsilon.$$

From a topological point of view, this can be established under some compactness conditions. Let us slightly rephrase our framework, writing

$$U = \mathcal{C}\times\mathcal{I}; \quad V = \mathcal{M}\times\mathcal{O};$$

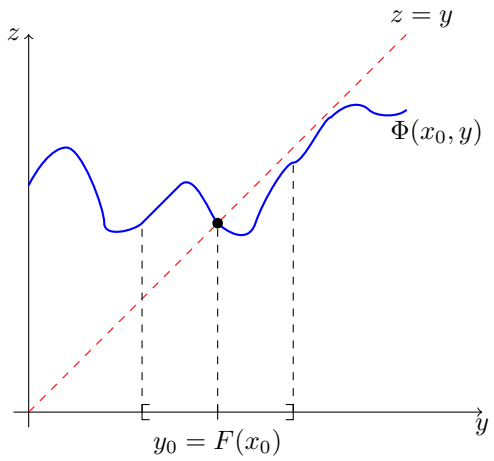


Figure 5.6: Approximate fixed points vs approximations of the exact fixed point. The intuition is as follows; assume that the fixed point equation $\Phi(x, y) = y$ has a solution operator $y = F(x)$ which is continuous. Then, in a compact neighborhood of (x, y), approximate fixed points are 'close' to the exact fixed point.

$$\Psi : U \times V \rightharpoonup \mathbb{R}_0^+ \text{ given by}$$
$$\Psi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M, \mathbf{u}^O) = d(\Phi(\mathbf{g}, \mathbf{u}^I, \mathbf{u}^M), (\mathbf{u}^M, \mathbf{u}^O)). \tag{5.10}$$

With this notation, well-posedness of the LGPAC ensures that there is a unique and continuous function $F : U \rightharpoonup V$ with some open domain $\mathrm{dom}(F) = U_0 \subseteq U$ such that $\Psi(u, F(u)) = 0$ for all $u \in U_0$. Then we have the following statement.

Lemma 5.3.3. *Let U, V be metric spaces and $\Psi : U \times V \rightharpoonup \mathbb{R}_0^+$ a continuous function. Let U_0, V_0 be compact subsets of U, V and suppose that the equation $\Psi(u, v) = 0$ has a unique solution $v \in V_0$ for each $u \in U_0$. Moreover, suppose that the solution $v = F(u)$ depends continuously on u. Then*

$$\text{for all } \epsilon > 0 \text{ there exists } \delta > 0 \text{ such that for all } u \in U_0, v \in V_0,\ \Psi(u, v) < \delta \Rightarrow d(F(u), v) < \epsilon. \tag{5.11}$$

Proof. First we prove a weaker version of (5.11), namely that

$$\text{for all } u \in U_0, \epsilon > 0, \text{ there exists } \delta > 0 \text{ such that for all } v \in V_0,\ \Psi(u, v) < \delta \Rightarrow d(F(u), v) < \epsilon. \tag{5.12}$$

We do this by contradiction. Assume that there exist $u \in U_0$ and $\epsilon > 0$ such that, for all $\delta > 0$ there exists $v \in V_0$ with

$$\Psi(u, v) < \delta \text{ but } d(F(u), v) \geq \epsilon;$$

consider a sequence $v_n \in V_0$ such that $\Psi(u, v_n) < \frac{1}{n}$ and $d(F(u), v_n) \geq \epsilon$. Since V_0 is (sequentially) compact, there is $v^* \in V_0$ and a converging subsequence $v_{n_k} \to v^*$. Then, by continuity of Ψ, we get $\Psi(u, v^*) = 0$, so that $v^* = F(u)$ (by uniqueness of solutions); however, by continuity of d, we get $d(F(u), v^*) \geq \epsilon$, which is a contradiction. This proves (5.12).

For the stronger version, let us fix $\epsilon > 0$. For each $u \in U_0$, let $\delta_1(u)$ be such that for any $v \in V_0$,

$$\Psi(u, v) < \delta_1(u) \Rightarrow d(F(u), v) < \epsilon/2.$$

Since F is continuous, for each $u \in U_0$ there is also $\delta_2(u)$ such that for any $u' \in U_0$,

$$d(u,u') < \delta_2(u) \Rightarrow d(F(u),F(u')) < \epsilon/2.$$

Next we prove the following claim: for each $u \in U_0$ there is $\delta_3(u)$ such that for any $u' \in U_0$ and $v \in V_0$,

$$d(u,u') < \delta_3(u) \Rightarrow d(\Psi(u,v),\Psi(u',v)) < \frac{\delta_1(u)}{2}; \tag{5.13}$$

notice how δ_3 depends on u only (i.e. is a uniform bound in $v \in V_0$). The claim can be proven by a compactness argument on V_0. Indeed, fix $u \in U_0$. Since Ψ is continuous, for each $v \in V_0$ there exist $\delta_4(u,v), \delta_5(u,v), \delta_6(u,v)$ such that for any $u' \in U_0$ and $v' \in V_0$,

$$d(u,u') < \delta_4(u,v) \text{ and } d(v,v') < \delta_5(u,v) \Rightarrow d(\Psi(u,v),\Psi(u',v')) < \frac{\delta_1(u)}{4};$$

$$d(v,v') < \delta_6(u,v) \Rightarrow d(\Psi(u,v),\Psi(u,v')) < \frac{\delta_1(u)}{4}.$$

Let $\delta_7(u,v) = \min(\delta_5(u,v),\delta_6(u,v))$ and cover V_0 with balls $V_0 = \bigcup_{v \in V_0} B_{\delta_7(u,v)}(v)$. By compactness, we can cover V_0 with only a finite number of balls $B_{\delta_7(u,v_i)}(v_i)$, $i = 1,\dots,N$.

Let $\delta_3(u) = \min_{1 \le i \le N} \delta_4(u,v_i)$, which will satisfy the claim (5.13). To see this, let $u' \in U_0$ be such that $d(u,u') < \delta_3(u)$. Let also $v \in V_0$ and take v_i such that $d(v,v_i) < \delta_7(u,v_i)$. Then

$$d(u,u') < \delta_4(u,v_i) \text{ and } d(v,v_i) < \delta_5(u,v_i) \text{ and } d(v,v_i) < \delta_6(u,v_i), \text{ thus}$$

$$d(\Psi(u,v_i),\Psi(u',v)) < \frac{\delta_1(u)}{4} \text{ and } d(\Psi(u,v),\Psi(u,v_i)) < \frac{\delta_1(u)}{4}, \text{ and so}$$

$$d(\Psi(u,v),\Psi(u',v)) < \frac{\delta_1(u)}{2};$$

this proves the claim.

Continuing the proof, let $\delta_8(u) = \min(\delta_2(u),\delta_3(u))$ and cover U_0 with balls $U_0 = \bigcup_{u \in U_0} B_{\delta_8(u)}(u)$. By compactness, we can cover U_0 with only a finite number of balls $B_{\delta_8(u_j)}(u_j)$, $j = 1,\dots,M$.

Let $\delta = \min_{1 \le j \le M} \frac{\delta_1(u_j)}{2}$, which will satisfy the desired property (5.11). To see this, let $u \in U_0$ and $v \in V_0$ be such that $\Psi(u,v) < \delta$. Take u_j such that $d(u,u_j) < \delta_8(u_j)$. We have that

$$d(u,u_j) < \delta_2(u_j) \text{ and } d(u,u_j) < \delta_3(u_j) \text{ and } \Psi(u,v) < \frac{\delta_1(u_j)}{2}, \text{ thus}$$

$$d(F(u),F(u_j)) < \epsilon/2 \text{ and } d(\Psi(u,v),\Psi(u_j,v)) < \frac{\delta_1(u_j)}{2} \text{ and } \Psi(u,v) < \frac{\delta_1(u_j)}{2}, \text{ so that}$$

$$d(F(u),F(u_j)) < \epsilon/2 \text{ and } \Psi(u_j,v) \le \Psi(u,v) + d(\Psi(u,v),\Psi(u_j,v)) < \delta_1(u_j), \text{ therefore}$$

$$d(F(u),F(u_j)) < \epsilon/2 \text{ and } d(F(u_j),v) < \epsilon/2, \text{ and so}$$

$$d(F(u),v) < \epsilon;$$

this concludes the proof. □

The above result shows that, in principle, solutions of the 'approximate fixed point problem' can provide approximations to solutions of the 'exact fixed point problem'. However, from the point of view of tracking computability, one would need to effectivize the proof of Lemma 5.3.3. Potentially, this may be doable by showing that, given modulus of continuity for F, Φ and Ψ, one could effectively produce a 'modulus of approximability' representative of the relation between ϵ and δ in (5.11). We would also require other additional effectivity properties with respect to the compactness assumptions (for example, a computable compact representation of U, V, and a dependence of the modulus of approximability on a representation of U_0, V_0). This could prove to be a quite cumbersome task and indeed we shall not pursue this direction. Our approach shall be to define a notion that captures this effective construction and use it as an assumption towards proving our main theorem.

Definition 5.3.4 (**Effective local reversibility and effective local behaviour**)**.** Let $\mathcal{G}$, Φ, U be as in Definition 5.3.1, with $\mathcal{G}$ well-posed on U. Let F be the specification of $\mathcal{G}$, having domain U. Let Ψ be as in (5.10). Let u_0, v_0 be computable elements of U and $F(U)$, respectively, and ϵ_1, ϵ_2 be computable reals. Let $U_0 = \bar{B}_{\epsilon_1}(u_0)$ and $V_0 = \bar{B}_{\epsilon_2}(v_0)$ be the corresponding *computable closed neighborhoods* of u_0, v_0. Suppose further that

- $U_0 \subseteq U = \mathrm{dom}(F)$;
- $F(U_0) \subseteq V_0$;
- there is an effective modulus of convergence $N : \mathbb{N} \to \mathbb{N}$ such that

$$\text{for all } \nu > 0, \text{ for all } u \in U_0, v \in V_0,\ \Psi(u, v) < 2^{-N(\nu)} \Rightarrow d(F(u), v) < 2^{-\nu}. \tag{5.14}$$

Under all these conditions, we say that the well-posedness of $\mathcal{G}$ is *effectively locally reversible* on U_0, or that F is *effectively locally well-behaved* on U_0.

Theorem 23 (**Tracking computability of LGPAC generable functions**)**.** *Let F be LGPAC-generable with domain U. Let $U_0 \subseteq U$ be a computable closed neighborhood such that F is effectively locally well-behaved on U_0. Then the restriction of F to U_0 is a tracking computable function.*

Proof. Given a code $c = \langle e, N\rangle$ of an element $u \in U_0$, we wish to construct a code $\varphi(c) = \langle e', N'\rangle$ of $F(u)$. Consider an LGPAC that generates F, with induced operator $\Phi : \mathcal{C} \times \mathcal{I} \times \mathcal{M} \rightharpoonup \tilde{\mathcal{M}} \times \mathcal{O}$. The code $\varphi(n)$ shall be constructed with N' being a code for the identity function and e' being a code for a sequence $\{e'\}(n)$ as follows. Let M be the effective modulus of convergence witnessing the local reversibility of F and let γ be the procedure $(n, \ell) \mapsto m$ that produces approximate fixed points of F, as in Lemma 5.3.2. Then we define $\{e'\}(n) = \gamma(c, M(n+1))$, which is an effective construction in n.

Clearly, the procedure $c \mapsto \langle e', N'\rangle$ is effective, so all we need to show is that this choice of e' is correct. For any code c of an element $u \in U_0$ (that is, $u = \bar{\alpha}_{\mathcal{C}\times\mathcal{I}}(c)$) and any $n \in \mathbb{N}$, we know that γ does terminate on input $(c, M(n+1))$. Let $v_n = \alpha_{\mathcal{M}\times\mathcal{O}}(\{e'\}(n)) = \alpha_{\mathcal{M}\times\mathcal{O}}(\gamma(c, M(n+1)))$. By construction of γ we know that $\Psi(u, v_n) < 2^{-M(n+1)}$. Then, by (5.14) we know that $d(F(u), v_n) < 2^{-n-1}$. Thus $v_n \to F(u)$. Moreover, v_n is a fast Cauchy sequence, for if $k_1, k_2 \geq n$, then

$$d(v_{k_1}, v_{k_2}) \leq d(F(u), v_{k_1}) + d(F(u), v_{k_2}) < 2^{-k_1-1} + 2^{-k_2-2} \leq 2^{-n}.$$

Therefore

$$\bar{\alpha}_{\mathcal{M}\times\mathcal{O}}(\varphi(c)) = \bar{\alpha}_{\mathcal{M}\times\mathcal{O}}(\langle e', N'\rangle) = \lim_{n\to\infty} \alpha(\{e'\}(n)) = \lim_{n\to\infty} v_n = F(u) = F(\bar{\alpha}_{\mathcal{C}\times\mathcal{I}}(c)),$$

so that φ is indeed a tracking function for F. We conclude that $F \restriction_{U_0}$ is tracking computable. □

5.4 Discussion

In this chapter we combined the two ideas of Chapters 3 and 4 into a more expressive GPAC with both the differential model and the continuous limit module (which we still call LGPAC). In order to ensure that the GPAC-generable functions be continuous, we had to make a slight technical change from the $\mathcal{X}$-GPAC presented in Chapter 3. In particular, we had to consider a version of the differential operator with domain consisting of continuously differentiable functions. In this way, well-posedness is restored as the necessary set of criteria for the GPAC semantics. One could adapt the definition of $\mathcal{X}$-GPAC from Chapter 3 to be consistent with the version in this chapter, but since the equational specification stays unchanged, the characterization would be the same as obtained in Theorem 17 (in terms of solutions to systems of PDAEs).

The main goal of this chapter was to connect our model with the notion of tracking computability, a paradigm of digital computation on the reals. We can state our original plan in terms of a conjecture as follows.

Conjecture A function is LGPAC-generable if and only if it is tracking computable.

Unfortunately, we were only able to prove one direction of this equivalence, namely that LGPAC-generable functions are tracking computable. Moreover, we had to introduce an extra assumption which we called effective local reversibility. This assumption allows us to use approximate fixed points to obtain approximations of the exact fixed point. The question of whether this condition can be relaxed remains an open problem. The main difficulty lies in finding an algorithm that produces approximations to the function generated by a well-posed GPAC. We can argue that this difficulty is a consequence of considering functions of more than one variable (and thus, essentially, dealing with systems of PDEs); in the case of functions of one variable (and systems of ODEs), standard results in analysis (e.g. the Picard-Lindelöf Theorem, [CL55]) allow us to consider iterative methods to obtain such fixed points.

The other direction, i.e. proving that tracking computable functions are LGPAC-generable, remains in the uncharted territory and a major open problem. We conjecture that this can be established and provide some ideas and suggestions for a possible approach in the next chapter.

Chapter 6

Conclusion and further work

In this thesis we studied a model of analog computation which combines the Shannon GPAC with the analog networks of Tucker and Zucker. The main difference from the models originally proposed by Shannon, Tucker and Zucker is that we consider streams carrying values on a general space $\mathcal{X}$. While there have been many attempts to generalize the Shannon GPAC to functions of more than one variable (notably Rubel's EAC [Rub93] and Bournez, Graça and Pouly's multidimensional GPAC [Pou15, BGP16]), the idea of changing the channel type is, to the best of our knowledge, novel with respect to the existing literature.

We have presented two different ways to increase the expressive power of the Shannon GPAC, by including a differential module (Chapter 3), a limit module (Chapter 4) or both (Chapter 5). We can represent these extensions as in Figure 6.1. In this chapter we shall address some open problems that are left to future research.

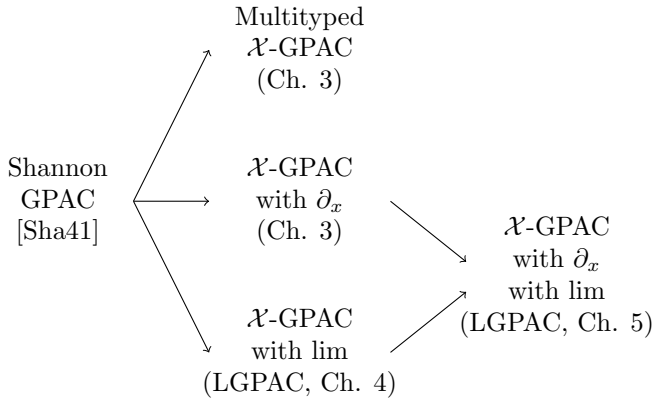


Figure 6.1: Different GPAC models; arrows indicate increase of expressive power.

6.1 Composition of functions

One can wonder if the two extensions presented in this thesis are comparable; say, could we derive the differential module from the limit module or vice versa? We now briefly address this

question. First of all, we can easily argue that the limit module cannot be derived from the other modules. The reason is that, as a consequence of Theorem 17, every $\mathcal{X}$-GPAC-generable function must satisfy a system of partial differential algebraic functions, and thus present some smoothness; in particular, $\mathcal{X}$-GPAC-generable functions are at least continuously differentiable. On the other hand, the limit module allows us to obtain non-differentiable functions. The most basic example is the *absolute value* function $x \mapsto |x|$. To see that the absolute value is LGPAC-generable, we first see that $\max(x, 0)$ can be expressed as a limit,

$$\lim_{t\to\infty} x\frac{\tanh(tx)+1}{2} = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x \leq 0 \end{cases} = \max(x, 0).$$

Since tanh is LGPAC-generable (in particular, it is differentially algebraic and thus Shannon GPAC-generable), we get that $\max(x, 0)$ can be otained using a continuous limit module (after finding a suitable rate of convergence T). Finally, the absolute value can simply be obtained as $|x| = \max(x, -x) = 2\max(x, 0) - x$.

As an intermediate step in the above reasoning, we have argued that $\max(x, 0)$ is LGPAC-generable. We can go one step further and consider a composition with a general function $u \in C(\mathbb{R})$, in order to conclude that $u \mapsto \max(u, 0)$ is an LGPAC-generable operation. Finally, if $u_1, u_2 \in C(\mathbb{R})$ are LGPAC-generable, then so is $\max(u_1, u_2)$, thanks to the relation $\max(u_1, u_2) = \max(u_1 - u_2, 0) + u_2$. Thus we get that $\max : \mathcal{X}^2 \to \mathcal{X}$ is another example of an LGPAC-generable operation which is not differentiable.

Let us now discuss the reverse direction, i.e. whether the differential module can be derived in the LGPAC model. Intuitively, we expect it to be so, since derivatives can be expressed as limits; in fact, the derivate is *defined* as the limit

$$\frac{du}{dx} = \lim_{h\to 0} \frac{u(x+h) - u(x)}{h}.$$

There are two obstacles in order to finish this line of thought. First, one has to convert $\lim_{h\to 0}$ into a limit of the form $\lim_{t\to\infty}$; this may be possible via a change of variables $h \mapsto t$. However one may need to make sure to capture both directional limits $h \to 0^-$ and $h \to 0^+$, otherwise one may get left and right derivatives instead. Second, one has to think about the term $u(x + h)$ appearing on the definition of derivative, and whether it can be obtained from the basic modules; in other words, is the operation $(t, u) \mapsto u(t + \cdot)$ LGPAC-generable? It may be necessary to introduce some type of composition module into this framework, and so we leave this task to further work.

6.2 Boundary value problems and eigenvalue problems

At the end of Chapter 3, we briefly discussed the possibility of considering general Sobolev spaces such as $\mathcal{X} = C^p(\Omega)$ or $\mathcal{X} = H^p(\Omega)$ in the definition of the $\mathcal{X}$-GPAC. We also motivated this idea by affirming that it could allow us to make interesting connections with the field of partial differential equations, where such spaces are ubiquitous. We now briefly expand on this point, by looking at other types of problems arising in the study of PDEs, and seeing how they could be applied to the GPAC model.

We begin with boundary value problems. In these problems, we are looking for a function u satisfying some behaviour in a domain Ω, and with prescribed *boundary data* on $\partial\Omega$. In this thesis, we have focused on time evolution problems (Definition 2.1.3) and therefore a possible approach consists in expressing boundary value problems in that framework.

Example 6.2.1 (**Poisson equation with boundary conditions**)**.** A typical boundary value problem is given by the Poisson equation. Let $\Omega \subseteq \mathbb{R}^n$ be a bounded domain with a suitably smooth boundary, say $\Omega = [0,1]^n$. Then we may wish to study the following problem: given functions f, g defined on Ω and $\partial\Omega$, find $u \in C^2(\Omega)$ such that

$$\begin{cases} \Delta u = f & \text{in } \Omega; \\ u = g & \text{on } \partial\Omega; \end{cases} \tag{6.1}$$

where $\Delta = \partial^2_{x_1} + \ldots + \partial^2_{x_n}$ denotes the Laplace operator.

We do not wish to provide a rigorous treatment of this equation, which is done extensively in standard textbooks such as [Eva98, Fol95]. In order to express solutions to this equation in the GPAC framework, we recall that the Poisson equation is used to describe a variety of physical phenomena; in particular, it appears in the study of heat conduction in describing steady-state solutions. In other words, a solution u to (6.1) can be obtained from a solution to the (non-homogeneous) heat equation

$$\begin{cases} \partial_t v = \Delta v - f & \text{in } \mathbb{T} \times \Omega; \\ v = g & \text{on } \mathbb{T} \times \partial\Omega; \\ v = 0 & \text{at } t = 0, \end{cases} \tag{6.2}$$

by taking $u = \lim\limits_{t\to\infty} v$. Since (6.2) describes a time evolution problem, it can in principle be tractable by the techniques presented in this thesis (see in particular Example 3.6.9 where we defined an $\mathcal{X}$-GPAC for the heat equation). Since in this case we are considering n spatial dimensions, we would have to adapt our $\mathcal{X}$-GPAC construction and consider n differential modules, computing $\partial_{x_1}, \ldots, \partial_{x_n}$. In any case, we can sketch a program to treat boundary value problems in our framework as follows:

- represent the solution of the boundary value problem as the steady-state of a time evolution problem;
- represent the solution of the time evolution problem as the specification of an $\mathcal{X}$-GPAC;
- obtain the solution of the boundary value problem by using a continuous limit module.

Example 6.2.2 (**Shooting method**)**.** Another possible approach for reducing a boundary value problem to an initial value problem, called the *shooting method*, is given as follows. Let $\Omega = [0,1]$ and consider the problem of finding $u \in C^2(\Omega)$ such that

$$u''(t) = f(t, u, u'), \quad u(0) = u_0, \quad u(1) = u_1. \tag{6.3}$$

The idea is to instead consider the initial value problem (depending on a parameter x)

$$u''(t) = f(t, u, u'), \quad u(0) = u_0, \quad u'(0) = x, \tag{6.4}$$

and find a suitable x such that $u(1) = u_1$. We can use a GPAC to generate solutions to (6.4). We also need to be able to compute the solution at the final time. Assuming u is continuous we get $u(1) = \lim_{t\to 1^-} u(t)$ and thus this can be achieved with a limit module. We can enforce the final condition with a feedback loop, repeating the technique from Chapter 3. Hence this approach may also be used to obtain solutions to one-dimensional boundary value problems.

Finally, we can look at eigenvalue problems; for example, suppose we have a linear operator $L : \mathcal{X} \rightharpoonup \mathcal{X}$. We wish to find values $\lambda \in \mathbb{R}$ such that there exists a non-zero $u \in \mathcal{X}$ with $Lu = \lambda u$. A typical example in PDEs concerns the study of the eigenvalues of the Laplacian, say in a bounded domain Ω,

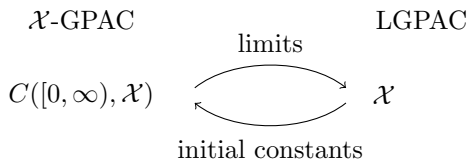


Figure 6.2: Recursive definition of a hierarchy of GPAC-generable functions.

$$\begin{cases} \Delta u = \lambda u & \text{in } \Omega; \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

In order to reason about eigenvalue problems, one can first consider them as boundary value problems or initial value problems, and then apply the abovementioned techniques. To deal with the parameter λ one can simply include it in the network as an additional channel. A technical obstacle concerns well-posedness, since solutions need not be unique; there may be multiple eigenvalues, and for each eigenvalue λ the eigenfunctions form a space of dimension at least one (and thus they are infinite). If the eigenvalues form a discrete set we can in principle narrow the search in a small compact neighborhood in order to have a unique eigenvalue. To deal with the eigenfunctions we can try to enforce, say, an extra condition on the norm of the solution (assuming $\mathcal{X}$ is a normed space). Of course, this would raise another question on how to compute norms with a GPAC. We leave this treatment to future work.

6.3 A hierarchy of LGPAC-generable functions

In Chapter 4 we introduced limit modules and arrived at a model which we called LGPAC. As we have seen, the limit module is an operation of type $C(\mathbb{T},\mathcal{X}) \to \mathcal{X}$ whose output is of 'one arity less' than the input. This can be seen as the reverse of the integrator module, whose initial constant $g \in \mathcal{X}$ is of 'one arity less' than the output in $C(\mathbb{T},\mathcal{X})$. This suggests one possible way of defining a hierarchy of 'computable functions' on $\mathcal{X}$, which we sketch as follows (see also Figure 6.2):

1. assume $\mathcal{X} = C(\mathbb{R})$ and take the subset $\mathcal{X}_0 \subseteq \mathcal{X}$ of Shannon GPAC-generable functions (ignore momentarily the fact that $C(\mathbb{R}) = C(\mathbb{R},\mathbb{R})$ is not the same as the class $C^1(\mathbb{T},\mathbb{R})$ appearing in the Shannon GPAC of Chapter 3); in this way, $\mathcal{X}_0$ corresponds to the class of differentially algebraic real functions as proven by Shannon and others;

2. using $\mathcal{X}_0$ as a class of 'valid initial constants' for integration in an $\mathcal{X}$-GPAC, define a class of $\mathcal{X}$-GPAC-generable functions in $C([0,\infty),\mathcal{X})$;

3. using continuous limit modules (that is, the LGPAC framework), define a subclass $\mathcal{X}_1 \subseteq \mathcal{X}$ of valid limits of the $\mathcal{X}$-GPAC-generable functions from the previous step; observe that this new class contains the gamma and Riemann zeta function, so $\mathcal{X}_1$ is strictly larger than $\mathcal{X}_0$;

4. the procedure can be iterated to get a hierarchy $\mathcal{X}_0 \subseteq \mathcal{X}_1 \subseteq \mathcal{X}_2 \subseteq \ldots \subseteq \mathcal{X}$ of 'computable functions' on $\mathcal{X}$.

We leave as an open problem the task of defining this hierarchy precisely and studying its properties. For example, deciding whether the hierarchy is closed under limits (we conjecture it is); whether it collapses at some level n; and whether the union $\bigcup_{n\in\mathbb{N}} \mathcal{X}_n$ has a simple characterization.

We would also hope that the union be different from $\mathcal{X}$, in order to have a non-trivial model of computation.

6.4 Equivalence with tracking computability

At the end of Chapter 5 we formulated the following conjecture relating LGPAC-generability and tracking computability

Conjecture A function is LGPAC-generable if and only if it is tracking computable.

There is a large volume of research and literature dedicated to the task of defining a continuous counterpart to the Church-Turing thesis, and an important step towards that goal consists in establishing equivalence results between various models of computation in continuous spaces. Looking at the above conjecture from that point of view, we consider this to be the most relevant open problem among those presented in this thesis. Some efforts were made in Chapter 5, specifically in the direction 'LGPAC-generable implies tracking computable', and we now provide some ideas and suggestions for a possible approach in the reverse direction.

The main idea is to somehow simulate the behavior of a Turing machine (or any other discrete model of computation) in an analog network. Indeed, there is some literature related to this task, such as the papers [BCGH07] and [CMC00]. In the first paper, the authors provide a way to represent: the state of a Turing machine computation (i.e. the machine state and the tape contents) as a real number; the transition function as a continuous function of type $\mathbb{R} \to \mathbb{R}$; and the discrete evolution of a Turing machine as the continuous evolution of a dynamical system. In the second paper, the authors show that a class of functions (which consists of GPAC-generable functions augmented with the functions θ_k from (6.5)) is closed under the same function constructs used to define primitive recursive functions. This class of functions consists of GPAC-generable functions augmented with the functions θ_k given by

$$\theta_k(x) = x^k \quad \text{for } x \geq 0; \quad \theta_k(x) = 0 \quad \text{for } x \leq 0, \tag{6.5}$$

which play a fundamental role in [CMC00] as functions which "*...check inequalities in a differentiable way, since* θ_k *is* $(k-1)$*-times differentiable...*"

In that paper, these are used to define special 'clock functions', which in turn are used to prove closure under iterations. With some care, their techniques may be adaptable to our framework.

We hope that in tackling these problems new insights can be acquired about the power of analog networks, and in particular the GPAC, as a model of analog computability.

Bibliography

[Abe65] Niels Henrik Abel. Solution de quelques problèmes à l'aide d'intégrales définies. In Ludwig Sylow and Sophus Lie, editors, *Oeuvres complètes d'Abel*, volume I, pages 11–27. Johnson, New York, 1965. Reprint of the Nouvelle éd., Christiania, 1881.

[BCGH06] Olivier Bournez, Manuel L. Campagnolo, Daniel S. Graça, and Emmanuel Hainry. The general purpose analog computer and computable analysis are two equivalent paradigms of analog computation. In Jin-yi Cai, S. Barry Cooper, and Angsheng Li, editors, *Theory and Applications of Models of Computation, Third International Conference, TAMC 2006, Beijing, China, May 15-20, 2006, Proceedings*, volume 3959 of *Lecture Notes in Computer Science*, pages 631–643. Springer, 2006.

[BCGH07] Olivier Bournez, Manuel L. Campagnolo, Daniel S. Graça, and Emmanuel Hainry. Polynomial differential equations compute all real computable functions on computable compact intervals. *Journal of Complexity*, 23(3):317–335, 2007.

[BGP16] Olivier Bournez, Daniel Graça, and Amaury Pouly. On the functions generated by the general purpose analog computer. submitted on 21 Jan 2016, arXiv:1602.00546, 2016.

[Bré11] Haïm Brézis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer-Verlag New York, 2011.

[Bus31] Vannevar Bush. The differential analyzer. a new machine for solving differential equations. *Journal of the Franklin Institute*, 212(4):447–488, 1931.

[Cau42] Augustin Cauchy. Mémoire sur l'emploi du calcul des limites dans l'intégration des équations aux dérivées partielles. *Comptes rendus hebdomadaires des séances*, 15:25–58, 1842.

[CH53] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 2. Interscience Publishers, Inc., 1953.

[CL55] Earl A. Coddington and Norman Levinson. *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[CMC00] Manuel Campagnolo, Cris Moore, and José Félix Costa. Iteration, inequalities, and differentiability in analog computers. *Journal of Complexity*, 16(4):642–660, 2000.

[CMC02] Manuel Campagnolo, Cris Moore, and José Félix Costa. An analog characterization of the Grzegorczyk hierarchy. *Journal of Complexity*, 18(4):977–1000, 2002.

[ES98] Yu V. Egorov and Mikhail A. Shubin. *Foundations of the classical theory of partial differential equations*. Springer-Verlag, 1998.

[Eva98] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, Rhode Island, 1998.

[Fol95] Gerald B. Folland. *Introduction to Partial Differential Equations*. Princeton University Press, 1995.

[GC03] Daniel Graça and José Félix Costa. Analog computers and recursive functions over the reals. *Journal of Complexity*, 19(5):644–664, 2003.

[Gra04] Daniel Graça. Some recent developments on shannon's general purpose analog computer. *Mathematical Logic Quarterly*, 50(4–5):473–485, 2004.

[Grz55] A. Grzegorczyk. Computable functions. *Fundamenta Mathematicae*, 42:168–202, 1955.

[Grz57] A. Grzegorczyk. On the defintions of computable real continuous functions. *Fundamenta Mathematicae*, 44:61–71, 1957.

[Had52] Jacques Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Dover, 1952.

[Har50] Douglas R. Hartree. *Calculating instruments and machines*. Cambridge University Press, 1950.

[Höl86] Otto Hölder. Ueber die eigenschaft der gammafunction keiner algebraischen differentialgleichung zu genügen. *Mathematische Annalen*, 28(1):1–13, 1886.

[Hol96] Per A. Holst. Svein Rosseland and the Oslo Analyzer. *IEEE Annals of the History of Computing*, 18(4):16–26, 1996.

[Jam12] Nick D. James. *Fixed Points in Analog Network Models*. PhD thesis, McMaster University, 2012.

[Joh96] Magnus Johansson. Early analog computers in Sweden - with examples from Chalmers University of Technology and the Swedish aerospace industry. *IEEE Annals of the History of Computing*, 18(4):27–33, 1996.

[JZ13] Nick D. James and Jeffery I. Zucker. A class of contracting stream operators. *The Computer Journal*, 56:15–33, 2013.

[KL03] Ulrich Kohlenbach and Branimir Lambov. Bounds on iterations of asymptotically quasi-nonexpansive mappings. *BRICS Report Series*, 10(51), 2003.

[KL10] Ulrich Kohlenbach and Laurentiu Leuştean. Asymptotically nonexpansive mappings in uniformly convex hyperbolic spaces. *Journal of the European Mathematical Society*, 12(1):71–92, 2010.

[Kle55] Stephen Kleene. Arithmetical predicates and function quantifiers. *Transactions of the American Mathematical Society*, 79:312–340, 1955.

[Ko91] Ker-I Ko. *Complexity Theory of Real Functions*. Birkäuser, 1991.

[Kow75] Sophie Kowalevski. Zur theorie der partiellen differentialgleichung. *Journal für die reine und angewandte Mathematik*, 80:1–32, 1875.

[Lac55a] D. Lacombe. Extension de la notion de fonction récursive aux fonctions d'une ou plusieurs variables réelles i. *Comptes Rendus des Séances d l'Académie des Sciences, Paris*, 240:2478–2480, 1955.

[Lac55b] D. Lacombe. Extension de la notion de fonction récursive aux fonctions d'une ou plusieurs variables réelles ii. *Comptes Rendus des Séances d l'Académie des Sciences, Paris*, 241:13–14, 1955.

[Lac55c] D. Lacombe. Extension de la notion de fonction récursive aux fonctions d'une ou plusieurs variables réelles iii. *Comptes Rendus des Séances d l'Académie des Sciences, Paris*, 241:151–153, 1955.

[LR87] Leonard Lipshitz and Lee Rubel. A differentially algebraic replacement theorem. *Proceedings of the American Mathematical Society*, 99(2):367–372, 1987.

[MC04] Jerzy Mycka and José Félix Costa. Real recursive functions and their hierarchy. *Journal of Complexity*, 20(6):835–857, 2004.

[Mil08] Jonathan W. Mills. The nature of the extended analog computer. *Physica D Nonlinear Phenomena*, 237(9):1235–1256, 2008.

[Moo96] Cris Moore. Recursion theory on the reals and continuous-time computation. *Theoretical Computer Science*, 162(1):23–44, 1996.

[OLBC10] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.

[Paz83] A. Pazy. *Semi-groups of linear operators and applications to partial differential equations*. Number 44 in Applied Math. Sciences. Springer, New York, 1983.

[PE74] Marian Pour-El. Abstract computability and its relations to the general purpose analog computer. *Transactions of the American Mathematical Society*, 199:1–28, 1974.

[PER79] Marian Pour-El and Ian Richards. A computable ordinary differential equation which possesses no computable solution. *Annals of Mathematical Logic*, 17:61–90, 1979.

[PER89] Marian Pour-El and Ian Richards. *Computability in Analysis and Physics*. Springer-Verlag, 1989.

[Pla20] Giovanni Antonio Amedeo Plana. Sur une nouvelle expression analytique des nombres bernoulliens, propre à exprimer en termes finis la formule générale pour la sommation des suites. *Mem. Accad. Sci. Torino*, 1(25):403–418, 1820.

[Pou15] Amaury Pouly. *Continuous models of computation: from computability to complexity*. PhD thesis, École Polytechnique and Universidade do Algarve, 2015.

[Rau91] Jeffrey Rauch. *Partial Differential Equations*, volume 128 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 1991.

[RR06] Michael Renardy and Robert C. Rogers. *An introduction to partial differential equations*, volume 13. Springer-Verlag New York, second edition, 2006.

[RS80] Michael Reed and Barry Simon. *Methods of modern mathematical physics: Functional analysis*. Academic Press, Inc, 1980.

[Rub93] Lee A. Rubel. The extended analog computer. *Advances in Applied Mathematics*, 14(1):39–50, 1993.

[Rud76] Walter Rudin. *Principles of Mathematical Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 3 edition, 1976.

[Rud91] Walter Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 2 edition, 1991.

[Sha41] Claude Shannon. Mathematical theory of the differential analyser. *Journal Mathematical Physics*, 20:337–354, 1941.

[SHT99] Viggo Stoltenberg-Hansen and John Tucker. Concrete models of computation for topological algebras. *Theoretical Computer Science*, 219:347–378, 1999.

[Sma93] James S. Small. General-purpose electronic analog computing: 1945-1965. *IEEE Annals of the History of Computing*, 15(2):8–18, 1993.

[TT80] William Thompson and Peter G. Tait. *Treatise on Natural Philosophy*, volume 1. Cambridge University Press, 2nd edition, 1880. Part I.

[Tur36] Alan Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936.

[TZ00] John V. Tucker and Jeffery I. Zucker. Computable functions and semicomputable sets on many sorted algebras. In Samson Abramsky, Dov Gabbay, and Tom Maibaum, editors, *Handbook of Logic for Computer Science*, volume V of *University Series in Mathematics*, pages 317–523. Oxford University Press, 2000.

[TZ04] John V. Tucker and Jeffery I. Zucker. Abstract versus concrete computation on metric partial algebras. *ACM Transactions on Computational Logic*, 5:611–668, 2004.

[TZ05] John V. Tucker and Jeffery I. Zucker. Computable total functions, algebraic specifications and dynamical systems. *Journal of Algebraic and Logic Programming*, 62:71–108, 2005.

[TZ07] John V. Tucker and Jeffery I. Zucker. Computability of analog networks. *Theoretical Computer Science*, 371:115–146, 2007.

[TZ11] John V. Tucker and Jeffery I. Zucker. Continuity of operators on continuous and discrete time streams. *Theoretical Computer Science*, 412:3378–3403, 2011.

[TZ14] John V. Tucker and Jeffery I. Zucker. Computability of operators on continuous and discrete time streams. *Computability*, 3:9–44, 2014.

[Wei00] Klaus Weihrauch. *Computable Analysis — An Introduction*. Texts in Theoretical Computer Science. Springer-Verlag Berlin Heidelberg, 2000.

Index