PhD Thesis

Neural Networks for Variational Problems in Engineering

Roberto López González

Director: Prof. Eugenio Oñate Ibañez de Navarra
Co-director: Dr. Eva Balsa Canto
Tutor: Dr. Lluís Belanche Muñoz

PhD Program in Artificial Intelligence
Department of Computer Languages and Systems
Technical University of Catalonia

21 September 2008

This page is intentionally left blank.

Neural Networks for Variational Problems in Engineering

This page is intentionally left blank.

Dedicado a mi familia,
Rufino, Esmeralda y Ana Belén.

This page is intentionally left blank.

Abstract

Many problems arising in science and engineering aim to find a function which is the optimal value of a specified functional. Some examples include optimal control, inverse analysis and optimal shape design. Only some of these, regarded as variational problems, can be solved analytically, and the only general technique is to approximate the solution using direct methods. Unfortunately, variational problems are very difficult to solve, and it becomes necessary to innovate in the field of numerical methods in order to overcome the difficulties.

The objective of this PhD Thesis is to develop a conceptual theory of neural networks from the perspective of functional analysis and variational calculus. Within this formulation, learning means to solve a variational problem by minimizing an objective functional associated to the neural network. The choice of the objective functional depends on the particular application. On the other side, its evaluation might need the integration of functions, ordinary differential equations or partial differential equations.

As it will be shown, neural networks are able to deal with a wide range of applications in mathematics and physics. More specifically, a variational formulation for the multilayer perceptron provides a direct method for solving variational problems. This includes typical applications such as function regression, pattern recognition or time series prediction, but also new ones such as optimal control, inverse problems and optimal shape design.

This addition of applications causes that a standard neural network is not able to deal with some particular problems, and it needs to be augmented. In this work an extended class of multilayer perceptron is developed which, besides the traditional neuron models and network architectures, includes independent parameters, boundary conditions and lower and upper bounds.

The computational performance of this numerical method is investigated here through the solution of different validation problems with analytical solution. Moreover, a variational formulation for an extended class of multilayer perceptron is applied to several engineering cases within optimal control, inverse problems or optimal shape design. Finally, this work comes with the open source neural networks C++ library Flood, which has been implemented following the functional analysis and calculus of variations theories.

This page is intentionally left blank.

Resumen

Muchos problemas en ciencia e ingenería se basan en encontrar una función para la cual un funcional dado toma un valor extremo. Algunos tipos de ejemplos son control óptimo, problemas inversos o diseño óptimo de formas. Sólo algunos de estos, conocidos como problemas variacionales, tienen solución analítica, y la única técnica general es aproximar la solución usando métodos directos. Sin embargo, los problemas variacionales son muy difíciles de resolver, y es necesario innovar en el campo de los métodos numéricos para paliar las dificultades.

El objetivo de esta Tesis Doctoral es desarrollar una teoría conceptual de las redes neuronales desde el punto de vista del análisis funcional y el cálculo de variaciones. Con esta formulación, aprender es equivalente a resolver un problema variacional minimizando un funcional objetivo asociado a la red neuronal. La elección del funcional objetivo depende de la aplicación en particular. Por otro lado, su evaluación puede necesitar la integración de funciones, ecuaciones diferenciales ordinarias o ecuaciones en derivadas parciales.

Como se verá, las redes neuronales son capaces de resolver una amplia gama de aplicaciones en matemáticas y física. Más específicamente, una formulación variacional para el perceptrón multicapa proporciona un método directo para la resolución de problemas variacionales. Esto incluye aplicaciones típicas como regresión de funciones, reconocimiento de patrones y predicción de series temporales, además de algunas nuevas, como control óptimo, problemas inversos y diseño óptimo de formas.

Dicho aumento de aplicaciones implica que una red neuronal estándar no sea apropiada o sea incapaz de tratar con algunos problemas particulares, y necesite ser ampliada. En este trabajo se desarrolla una clase extendida de perceptrón multicapa que, además de los modelos de neurona y las arquitecturas de red habituales, incluye parámetros independientes, condiciones de contorno y cotas inferiores y superiores.

El rendimiento computacional de este método numérico es investigado mediante la resolución de diversos problemas de validación con solución analítica. Además, una formulación variacional para una clase extendida de perceptrón multicapa es aplicada a distintos casos en ingeniería dentro del control óptimo, los problemas inversos y el diseño óptimo de formas. Finalmente, este trabajo se acompaña de la librería C++ en código abierto de redes neuronales llamada Flood, que ha sido implementada bajo la teoría del análisis funcional y el cálculo de variaciones.

This page is intentionally left blank.

Acknowledgments

Special thanks to my Director, Prof. Eugenio Oñate, for his belief and support in this project. Many thanks to my Co-Director, Dr. Eva Balsa, and my Tutor, Dr. Lluìs Belanche, for their fruitful comments and suggestions.

I also express my gratitude to all the people who has collaborated to any part of this PhD Thesis. They include Dr. Carlos Agelet, Begoña Carmona, Dr. Michele Chiumenti, Dr. Pooyan Dadvand, Alan Daring, Xavier Diego, Enrique Escolano, Dr. Roberto Flores, Dr. Julio Garcia, Kevin Lau and Inma Ortigosa.

Finally I would like to thank all my colleagues at the International Center for Numerical Methods in Engineering (CIMNE) for their pleasant friendship.

This page is intentionally left blank.

Contents

List of Figures

List of Tables

This page is intentionally left blank.

List of Applications

Applications listed by class

1. Function regression:
    - Residuary resistance of sailing yachts modeling, see Section 4.2.
    - Airfoil self-noise prediction, see Section 4.3.
2. Optimal control:
    - Fed batch fermenter, see Section 6.3.
    - Aircraft landing systems, see Section 6.4.
3. Inverse problems:
    - Microstructural modeling of aluminium alloys, see Section 7.3.
4. Optimal shape design:
    - Optimal airfoil design, see Section 8.3.

Applications listed by area

1. Aeronautical:
    - Airfoil self-noise prediction, see Section 4.3
    - Optimal control of aircraft landing systems, see Section 6.4.
    - Optimal airfoil design, see Section 8.3.
2. Chemical:
    - Maximum yield in a fed batch fermenter, see Section 6.3.
3. Naval:
    - Residuary resistance of sailing yachts prediction, see Section 4.2.
4. Metallurgical:
    - Microstructural modeling of aluminium alloys, see Section 7.3.

This page is intentionally left blank.

Chapter 1

Introduction

Queen Dido of Carthage was apparently the first person to attack a problem that can readily be solved by using the calculus of variations. Dido, having been promised all of the land she could enclose with a bull's hide, cleverly cut the hide into many lengths and tied the ends together. Having done this, her problem was to find the closed curve with a fixed perimeter that encloses the maximum area [53]. The problem is based on a passage from Virgil's Aeneid:

The Kingdom you see is Carthage, the Tyrians, the town of Agenor;
But the country around is Libya, no folk to meet in war.
Dido, who left the city of Tyre to escape her brother,
Rules here-a long a labyrinthine tale of wrong
Is hers, but I will touch on its salient points in order...
Dido, in great disquiet, organised her friends for escape.
They met together, all those who harshly hated the tyrant
Or keenly feared him: they seized some ships which chanced to be ready...
They came to this spot, where to-day you can behold the mighty
Battlements and the rising citadel of New Carthage,
And purchased a site, which was named 'Bull's Hide' after the bargain
By which they should get as much land as they could enclose with a bull's hide.

Despite the circle appears to be an obvious solution to Dido's problem, proving this fact is rather difficult. Zenodorus proved that the area of the circle is larger than that of any polygon having the same perimeter, but the problem was not rigorously solved until 1838 by Jakob Steiner [99].

The history of variational calculus dates back to the ancient Greeks, but it was not until the seventeenth century in western Europe that substantial progress was made. A problem of historical interest is the brachistochrone problem, posed by Johann Bernoulli in 1696. The term brachistochrone derives from the Greek 'brachistos' (the shortest) and 'chronos' (time):

Given two points A and B in a vertical plane, what is the curve traced out by a particle acted on only by gravity, which starts at A and reaches B in the shortest time?

Sir Isaac Newton was challenged to solve the problem, and did so the very next day. In fact, the solution to the brachistochrone problem, which is a segment of a cycloid, is credited to Johann and Jacob Bernoulli, Sir Isaac Newton and Guillaume de L'Hôpital [53].

In that context, the so called Dido's problem was renamed as the isoperimetric problem, and it was stated as:

Of all simple closed curves in the plane of a given length, which encloses the maximum area?

Another important variational problem is the data modeling problem. A conventional approach is the method of least squares, which was first proposed by Adrien Marie Legendre and Carl Friedrich Gauss in the early nineteenth century as a way of inferring planetary trajectories from noisy data [6].

Note that, in all these problems, curves are sought which are optimal in some sense. More specifically, the aim of a variational problem is to find a function which is the minimal or the maximal value of a specified functional. By a functional, we mean a correspondence which assigns a number to each function belonging to some class [29]. The calculus of variations gives methods for finding extremals of functionals, and problems that consist in finding minimal and maximal values of functionals are called variational problems [29].

From an engineering point of view, variational problems can be classified according to the way in which they can be applied for a particular purpose. In this way, some classes of variational problems of practical interest are optimal control, inverse or optimal shape design. They are often defined by integrals, ordinary differential equations or partial differential equations.

Optimal control is playing an increasingly important role in the design of modern engineering systems. The aim here is the optimization, in some defined sense, of a physical process. More specifically, the objective of these problems is to determine the control signals that will cause a process to satisfy the physical constraints and at the same time minimize or maximize some performance criterion [53] [10]. As a simple example, consider the problem of a rocket launching a satellite into an orbit around the earth. An associated optimal control problem is to choose the controls (the thrust attitude angle and the rate of emission of the exhaust gases) so that the rocket takes the satellite into its prescribed orbit with minimum expenditure of fuel or in minimum time.

Inverse problems can be described as being opposed to direct problems. In a direct problem the cause is given, and the effect is determined. In an inverse problem the effect is given, and the cause is estimated [54] [90] [85]. There are two main types of inverse problems: input estimation, in which the system properties and output are known and the input is to be estimated; and properties estimation, in which the the system input and output are known and the properties are to be estimated. Inverse problems can be found in many areas of science and engineering. A typical inverse problem in geophysics is to find the subsurface inhomogeneities from collected scattered fields caused by acoustic waves sent at the surface.

Optimal shape design is a very interesting field for industrial applications. The goal in these problems is to computerize the development process of some tool, and therefore shorten the time it takes to create or to improve the existing one. Being more precise, in an optimal shape design process one wishes to optimize some performance criterium involving the solution of a mathematical model with respect to its domain of definition [18]. One example is the design of airfoils, which proceeds from a knowledge of the boundary layer properties and the relation between geometry and pressure distribution [31] [74]. The performance goal here might vary: weight reduction, stress reinforcement, drag reduction and even noise reduction can be obtained. On the other hand, the airfoil may be required to achieve this performance with constraints on thickness, pitching moment, etc.

However, while some simple variational problems can be solved analytically by means of the Euler-Lagrange equation [29], the only practical technique for general problems is to approximate the solution using direct methods [29]. The fundamental idea underlying the so called direct methods is to consider the variational problem as a limit problem for some function optimization

problem in many dimensions [29]. Two direct methods of considerable concern are those due to Euler and Ritz [29]. Alternative techniques based on Laguerre polynomials [19], Legendre polynomials [22], Chebyshev polynomials [48], or more recently wavelets [86] [51], for instance, have been also proposed. Unfortunately, all these problems are difficult to solve. Deficient approximation or convergence properties of the base functions, the usually large dimension in the resulting function optimization problem, the presence of local minima or solutions presenting oscillatory behaviors are some of the most typical complications. Therefore new numerical methods need to be developed in order to overcome that troubles.

Unfortunately, variational problems might be extremely difficult to be solved. Deficient approximation or convergence properties of the base functions, the usually large dimension in the resulting function optimization problem, the presence of local minima or solutions presenting oscillatory behaviors are some of the most typical complications. Therefore, more effort is needed in order to surmount these difficulties.

During its development, artificial intelligence has been moving toward new methods of knowledge representation and processing that are closer to human reasoning. In this regard, a new computational paradigm has been established with many developments and applications - artificial neural networks. An artificial neural network, or simply a neural network, can be defined as a biologically inspired computational model which consists of a network architecture composed of artificial neurons [41]. This structure contains a set of parameters, which can be adjusted to perform certain tasks.

Even though neural networks have similarities to the human brain, they are not meant to model it, but to be useful models for problem-solving and knowledge-engineering in a 'humanlike' way. The human brain is much more complex and unfortunately, many of its cognitive functions are still not well known. But the more about the human brain is learnt, the better computational models can be developed and put to practical use.

One way to understand the ideas behind neural computation is to look at the history of this science [3]. McCulloch and Pitts described in 1943 a model of a neuron that is binary and has a fixed threshold [72]. A network of such neurons can perform logical operations, and it is capable of universal computation. Hebb, in his book published in 1949 [42], proposed neural network architectures and the first training algorithm. This is used to form a theory of how collections of cells might form a concept. Rosenblatt, in 1958, put together the ideas of McCulloch, Pitts and Hebb to present the perceptron neuron model and its training algorithm [87]. Minsky and Papert demonstrated the theoretical limits of the perceptron in 1969 [73]. Many researchers then abandoned neural networks and started to develop other artificial intelligence methods and systems.

New models, among them the associative memories [47], self-organizing networks [55], the multilayer perceptron and the back-propagation training algorithm [89], or the adaptive resonance theory (ART) [20] were developed later, which brought researchers back to the field of neural networks. Now, many more types of neural networks have been designed and used. The bidirectional associative memory, radial basis functions, probabilistic RAM neural networks, fuzzy neural networks and oscillatory neural networks are only a small number of models developed.

From all that kinds of neural network, the multilayer perceptron is a very important one, and much of the literature in the area is referred to it. Traditionally, the learning problem for the multilayer perceptron has been formulated in terms of the minimization of an error function of the free parameters, in order to fit the neural network outputs to an input-target data set [15]. In that way, the learning tasks allowed are data modeling type problems, such as function regression, pattern recognition or time series prediction.

The function regression problem can be regarded as the problem of approximating a function from noisy or corrupted data [41]. Here the neural network learns from knowledge represented by

a training data set consisting of input-target examples. The targets are a specification of what the response to the inputs should be [15].

The pattern recognition (or classification) problem can be stated as the process whereby a received pattern, characterized by a distinct set of features, is assigned to one of a prescribed number of classes [41]. Here the neural network learns from knowledge represented by a training data set consisting of input-target examples. The inputs include a set of features which characterize a pattern. The targets specify the class that each pattern belongs to [15].

The time series prediction problem can be regarded as the problem of predicting the future state of a system from a set of past observations to that system [41]. A time series prediction problem can also be solved by approximating a function from input-target data, where the inputs include data from past observations and the targets include corresponding future observations in relation to that inputs [15].

Regarding engineering applications, neural networks are nowadays being used to solve a whole range of problems [58]. Most common case studies include modeling hitherto intractable processes, designing complex feed-back control signals, building meta-models for posterior optimization or detecting device faults. All that applications need from an input-target data set in order to evaluate the error function, and therefore fall into one of the three categories of data modeling enumerated above.

In this work a variational formulation for the multilayer perceptron is presented. Within this formulation, the learning problem lies in terms of solving a variational problem by minimizing an objective functional of the function space spanned by the neural network. The choice of a suitable objective functional depends on the particular application and it might need the integration of functions, ordinary differential equations or partial differential equations in order to be evaluated. As we will see, neural networks are not only able to solve data modeling applications, but also a wide range of mathematical and physical problems. More specifically, a variational formulation for neural networks provides a direct method for solving variational problems.

In order to validate this numerical method we train a multilayer perceptron to solve some classical problems in the calculus of variations, and compare the neural network results against the analytical result. This variational formulation is also applied to some traditional learning tasks, such as function regression. Finally validation examples and real applications of different classes of variational problems in engineering, such as optimal control, inverse analysis and shape design are also included.

It is important to mention that some efforts have been already performed in this field. Zoppoli et al. [103] discussed the approximation properties of different classes of neural networks as linear combinations of non fixed basis functions, and applied this theory to some stochastic functional optimization problems. In [91], Sarkar and Modak make use of a neural network to determine several optimal control profiles for different chemical reactors. Also, Franco-Lara and Weuster-Botz [35] estimated optimal feeding strategies for bed-batch bioprocesses using neural networks.

This work complements that research and introduces a number of novel issues. First, a conceptual theory for neural networks from a variational point of view is written. In particular, we introduce the idea of the function space spanned by a multilayer perceptron. This neural network can also be extended so as to include independent parameters, boundary conditions and bounds. On the other hand, learning tasks are here stated as variational problems. They are defined by an objective functional, which has a parameterized function associated to it. In addition, we explain the solution approach of the reduced function optimization problem by means of the training algorithm. Second, the range of engineering applications for neural networks is here augmented so as to include shape design, optimal control and inverse problems. The performance of this computational tool is studied through the solution of a validation case for each of that type of problems.

Third, the embed of different numerical methods, such as the Simpson method, the Runge-Kutta method or the Finite Element Method for evaluating how well a neural network does an activity is here considered.

In summary, the emphasis of this PhD Thesis is placed on the demonstration of applicability of the multilayer perceptron to various classes of variational problems in engineering. In this regard, a single software tool can be used for many different applications, saving time and effort to the engineers and therefore reducing cost to the companies.

All the examples presented here are written in a self-contained fashion, and the main concepts are described for every single example, so that the reader needs only to look at the type of applications which are interesting for him.

Together with this PhD Thesis there is the open source neural networks C++ library Flood [61]. It uses the same concepts and terminology as in here, and contains everything needed to solve most of the applications included.

In Chapter 2 the basic mathematical theory concerning the calculus of variations is presented. These are the most important concepts from which the rest of this work is developed.

In Chapter 3 an extended class of multilayer perceptron is presented so as to include independent parameters, boundary conditions and lower and upper bounds. The learning problem for that neural network is then formulated from the perspective of functional analysis and variational calculus.

In Chapter 4 the task of data modeling is described from a variational point of view. In this way, function regression and pattern recognition fit in the variational formulation proposed. Several practical applications are also solved.

In Chapter 5 some classical problems in the calculus of variations are solved using neural networks. They are intended to be used for validation purposes, but also as a starting point to understand the solution approach of more complex applications.

In Chapter 6 the most important aspects of optimal control are reviewed and this type of problems is formulated as a learning task for the multilayer perceptron. In order to validate this direct method an optimal control problem with analytical solution is solved. Two other real applications are also approached with a neural network.

In Chapter 7 the inverse problems theory is introduced and stated as a possible application for neural networks. This is also validated through two different case studies artificially generated.

In Chapter 8 the mathematical basis of optimal shape design is stated and the suitability for a multilayer perceptron to solve this problems enunciated. Validation examples and more complex applications are also included here.

In Annex A the software model of Flood [61] is constructed following a top-down approach. This provides different views of this class library, from the highest conceptual level to the details.

In Annex B, and for the sake of completeness, some basics of numerical methods for integration of functions, ordinary differential equations and partial differential equations are included. They are meant to be utilities to be used by the neural network when needed, in order to evaluate the objective functional.

This page is intentionally left blank.

Chapter 2

Preliminaries

The calculus of variations, also known as functional differentiation, is a branch of mathematics which gives methods for finding minimal and maximal values of functionals. In this way, problems that consist in finding extrema of functionals are called variational problems.

In this Chapter we introduce the main concepts concerning optimum values of functionals. We also cite some methods to solve problems in the calculus of variations.

2.1 Extreme values of functionals

Variational problems involve determining a function for which a specific functional takes on a minimum or a maximum value. Here we introduce some concepts concerning functionals, their variations and their extreme values.

Vector spaces

To begin, let us introduce the concept of vector space.

Definition 1 (Vector space)**.** A vector space V is a set that is closed under the operations of element addition and scalar multiplication. That is, for any two elements $\mathbf{u}, \mathbf{v} \in V$ and any scalar $\alpha \in \mathbb{R}$

1. $\mathbf{u} + \mathbf{v} \in V$.
2. $\alpha\mathbf{v} \in V$.

Example 1. The Euclidean n-space, denoted $\mathbb{R}^n$, and consisting of the space of all n-tuples of real numbers, $(u_1, u_2, \ldots, u_n)$, is a vector space. Elements of $\mathbb{R}^n$ are called n-vectors. The operation of element addition is componentwise; the operation of scalar multiplication is multiplication on each term separately.

Normed vector spaces

In a vector space, besides the operations of element addition and scalar multiplication, usually there is introduced a norm [56].

Definition 2 (Normed vector space)**.** A vector space V is said to be normed if each element $\mathbf{u} \in V$ is assigned a nonnegative number $\|\mathbf{u}\|$, called the norm of $\mathbf{u}$, such that

1. $\|\mathbf{u}\| = 0$ if and only if $\mathbf{u} = \mathbf{0}$.
2. $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|$.
3. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Example 2. The Euclidean n-space is a normed vector space. A norm of a n-vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ can be

$$\|\mathbf{u}\| = \sqrt{\sum_{i=1}^{n} u_i^2}. \tag{2.1}$$

Function spaces

Here we introduce the concept of function space, which is directly related to the concept of normed vector space [56].

Definition 3 (Function space)**.** A function space is a normed vector space whose elements are functions. Function spaces might be of infinite dimension.

Example 3. The space $C^n(x_a, x_b)$, consisting of all functions $y(x)$ defined on a closed interval $[x_a, x_b]$ which have bounded continuous derivatives up to order n, is a function space of infinite dimensions. By addition of elements in C^n and multiplication of elements in C^n by scalars, we mean ordinary addition of functions and multiplication of functions by numbers. A norm in $C^n(x_a, x_b)$ can be

$$\|y(x)\|_n = \sum_{i=0}^{n} \sup_{x\in[x_a,x_b]} |y^{(i)}(x)|, \tag{2.2}$$

where $y^{(0)}(x)$ denotes the function $y(x)$ itself and $y^{(i)}(x)$ its derivative of order i. Thus, two functions in C^n are regarded as close together if the values of the functions themselves and of all their derivatives up to order n are close together. It is easily verified that all the axioms of a normed linear space are satisfied for the space C^n.

Two important C^n function spaces are C^0, the space of continuous functions, and C^1, the space of continuously differentiable functions.

Example 4. The space $P^n(x_a, x_b)$, consisting of all polynomials of order n defined on an interval $[x_a, x_b]$, is a function space of dimension $n + 1$. Elements in P^n are of the form

$$p_n(x) = \sum_{k=0}^{n} \alpha_k x^k. \tag{2.3}$$

The operations of element addition and scalar multiplication in $P^n(x_a, x_b)$ are defined just as in Example 3. The l-norm is defined as

$$\|P^n\|_l = \left(\sum_{k=0}^{n} |\alpha_k|^l \right)^{1/l}, \tag{2.4}$$

for $l \geq 1$. This formula gives the special cases

$$\|p_n\|_1 = \sum_{k=0}^{n} |\alpha_k|, \tag{2.5}$$

$$\|p_n\|_2 = \sqrt{\sum_{k=0}^{n} |\alpha_k|^2}, \tag{2.6}$$

$$\|p_n\|_\infty = \max_{0 \leq k \leq n} (|\alpha_k|). \tag{2.7}$$

It is easily verified that all the axioms of a normed linear space are satisfied for the space P^n.

Functionals

By a functional, we mean a correspondence which assigns a number to each function belonging to some class [56].

Definition 4 (Functional)**.** Let V be some function space. A functional $F[y(x)]$ is a correspondence which assigns a number $F \in \mathbb{R}$ to each function $y(x) \in V$

$$\begin{array}{rccc} F: & V & \to & \mathbb{R} \\ & y(x) & \mapsto & F[y(x)] \end{array}$$

V is called the domain of the functional.

Example 5 (Arc length)**.** Let $A = (x_a, y_a)$ and $B = (x_b, y_b)$ be two points on the plane and consider the collection of all functions $y(x) \in C^1(x_a, x_b)$ which connect A to B, i.e., such that $y(x_a) = y_a$ and $y(x_b) = y_b$. The arc-length L of a curve $y(x)$ is a functional. The value $L[y(x)]$ is given by the integral

$$\begin{aligned} L[y(x)] &= \int_A^B ds \\ &= \int_A^B \sqrt{dx^2 + dy^2} \\ &= \int_{x_a}^{x_b} \sqrt{1 + [y'(x)]^2} dx. \end{aligned} \tag{2.8}$$

Figure 2.1 represents graphically the arc length functional.

Example 6 (Sum squared error)**.** Let $y(x)$ be a given function, and consider a collection of data points $(x_1, y_1), \ldots, (x_n, y_n)$. The squared error E of the curve $y(x)$ with respect to the points $(x_1, y_1), \ldots, (x_n, y_n)$ is a functional. The value $E[y(x)]$ is given by

$$E[y(x)] = \sum_{i=1}^{n} \left(y(x_i) - y_i\right)^2. \tag{2.9}$$

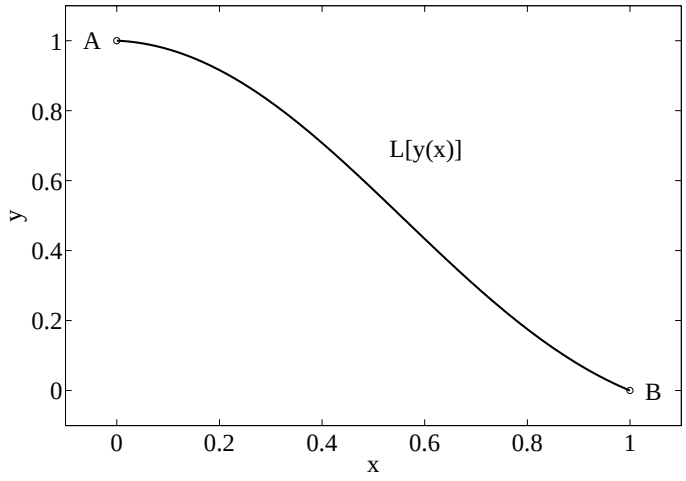


Figure 2.1: Illustration of the arc length functional.

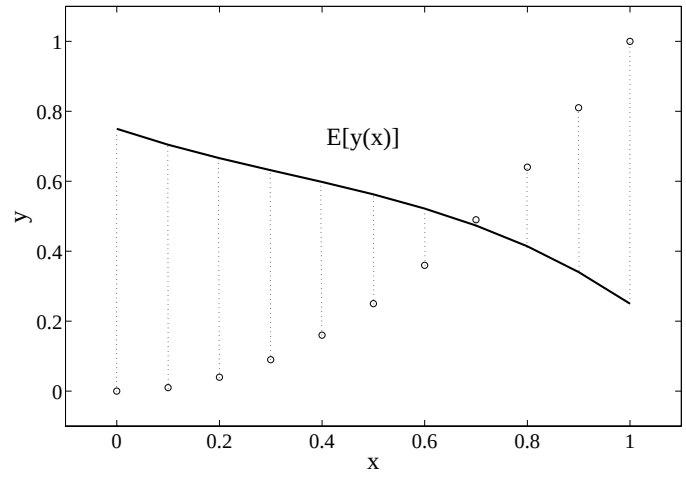


Figure 2.2: Illustration of the sum squared error functional.

Figure 2.1 shows the sum squared error functional.

Both of these examples have one property in common that is a characteristic feature of all functionals: Given a functional F, to each function $y(x)$ there corresponds a unique number $F[y(x)]$, just as when we have a function y, to each number x there corresponds a unique number $y(x)$.

Continuity of functionals

The concept of continuity of functionals naturally appears in functional analysis [56].

Definition 5 (Continuous functional)**.** Let V be some function space. A functional $F[y(x)]$ defined on V is said to be continuous at $y(x) \in V$ if for any $\epsilon > 0$ there is a $\delta > 0$ such that, if for some $\hat{y}(x) \in V$

$$\|y(x) - \hat{y}(x)\| < \delta \tag{2.10}$$

then

$$|F[y(x)] - F[\hat{y}(x)]| < \epsilon. \tag{2.11}$$

In the sequel we shall consider only continuous functionals and, for brevity, we shall omit the word 'continuous'.

Linearity of functionals

Let us introduce the concept of linearity of functionals, which will be useful to us later [56].

Definition 6 (Linear functional)**.** Let V be some function space. A functional $F[y(x)]$ defined on V is said to be linear if

1. $F[\alpha y(x)] = \alpha F[y(x)]$ for any $\alpha \in \mathbb{R}$ and any $y(x) \in V$.
2. $F[y_1(x) + y_2(x)] = F[y_1(x)] + F[y_2(x)]$ for any $y_1(x), y_2(x) \in V$.
3. $F[y(x)]$ is continuous for any $y(x) \in V$.

The increment of a functional

In order to consider the variation of a functional, we must define first the concept of increment [53].

Definition 7 (Increment of a functional)**.** Let V be some function space and let F be a functional defined on V. The increment of F, denoted by ΔF, is defined as

$$\Delta F(y(x), \delta y(x)) = F[y(x) + \delta y(x)] - F[y(x)], \tag{2.12}$$

for any $y(x), \delta y(x) \in V$.

The variation of a functional

We are now ready for considering the variation (or differential) of a functional [53].

Definition 8 (Variation of a functional)**.** Let V be some function space and let F be a functional defined on V. Let write the increment of F in the form

$$\Delta F[y(x), \delta y(x)] = \delta F[y(x), \delta y(x)] + G[y(x), \delta y(x)] \cdot \|\delta y(x)\|, \tag{2.13}$$

where $y(x), \delta y(x) \in V$ and δF is a linear functional. If

$$\lim_{\|\delta y(x)\| \to 0} G[y(x), \delta y(x)] = 0, \tag{2.14}$$

then F is said to be differentiable on $y(x)$ and δF is called the variation (or differential) of F at $y(x)$.

Extrema of functionals

Next we use the concept of variation to establish a necessary condition for a functional to have an extremum. We begin by introducing the concept of extrema of functionals [53].

Definition 9 (Extremum of a functional)**.** Let V be some function space and let F be a functional defined on V. The function $y^*(x)$ is said to yield a relative minimum (maximum) for F if there is an $\epsilon > 0$ such that

$$F[y(x)] - F[y^*(x)] \geq 0 \quad (\leq 0), \tag{2.15}$$

for all $y(x) \in V$ for which $\|y(x) - y^*(x)\| < \epsilon$.

If Equation (2.15) is satisfied for arbitrarily large ϵ, then $F[y^*(x)]$ is a global minimum (maximum). The function $y^*(x)$ is called an extremal, and $F[y^*(x)]$ is referred to as an extremum.

The fundamental theorem in the calculus of variations provides the necessary condition for a function to be an extremal of a functional [53].

Theorem 1 (The fundamental theorem in the calculus of variations)**.** Let V be some function space and let F be a functional defined on V. If $y^*(x) \in V$ is an extremal of F then

$$\delta F[y^*(x), \delta y(x)] = 0, \tag{2.16}$$

for any $\delta y(x) \in V$.

2.2 The simplest variational problem

We shall now consider what might be called the 'simplest' variational problem [36], which can be formulated as follows:

Problem 1 (The simplest variational problem)**.** Let V be the space of all smooth functions $y(x)$ defined on an interval $[x_a, x_b]$ which satisfy the boundary conditions $y(x_a) = y_a$ and $y(x_b) = y_b$. Find a function $y^*(x) \in V$ for which the functional

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx, \tag{2.17}$$

defined on V, takes on a minimum or maximum value.

In other words, the simplest variational problem consists of finding an extremum of a functional of the form (2.17), where the class of admissible functions consists of all smooth curves joining two points. Such variational problems can be solved analytically by means of the Euler-Lagrange Equation [36].

Theorem 2 (Euler-Lagrange equation)**.** Let V be the space of all smooth functions $y(x)$ defined on an interval $[x_a, x_b]$ which satisfy the boundary conditions $y(x_a) = y_a$ and $y(x_b) = y_b$, and let $F[y(x)]$ be a functional defined on V of the form

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx \tag{2.18}$$

Then, a necessary condition for the function $y^*(x) \in V$ to be an extremum of $F[y(x)]$ is that it satisfies the differential equation

$$\frac{\partial \widehat{F}}{\partial y} - \frac{d}{dx}\frac{\partial \widehat{F}}{\partial y'} = 0. \tag{2.19}$$

The Euler-Lagrange equation plays a fundamental role in the calculus of variations and it is, in general, a second order differential equation. The solution then will depend on two arbitrary constants, which can be determined from the boundary conditions $y(x_a) = y_a$ and $y(x_b) = y_b$.

Example 7 (The brachistochrone problem)**.** The statement of this problem is:

Given two points $A = (x_a, y_a)$ and $B = (x_b, y_b)$ in a vertical plane, what is the curve traced out by a particle acted on only by gravity, which starts at A and reaches B in the shortest time?

Figure 7 is a graphical statement of the brachistochrone problem.

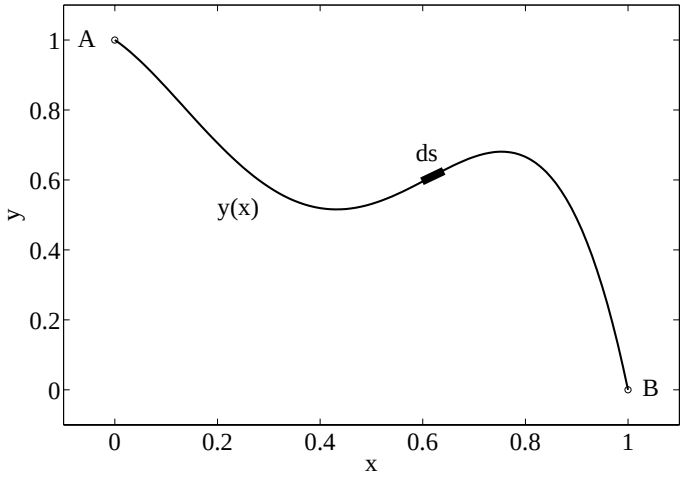


Figure 2.3: The brachistochrone problem statement.

The time to travel from point A to point B is given by the integral

$$t = \int_A^B \frac{ds}{v(s)}, \tag{2.20}$$

where s is the arc length and v is the speed. The speed at any point is given by a simple application of conservation of energy,

$$mgy_a = mgy + \frac{1}{2}mv^2, \tag{2.21}$$

where $g = 9.81$ is the gravitational acceleration. This equation gives

$$v = \sqrt{2g(y_a - y)}. \tag{2.22}$$

Plugging this into Equation (2.20), together with the identity

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2} \\ &= \sqrt{1 + [y'(x)]^2}dx, \end{aligned} \tag{2.23}$$

yields

$$t = \frac{1}{\sqrt{2g}} \int_{x_a}^{x_b} \sqrt{\frac{1 + [y'(x)]^2}{y_a - y(x)}}dx. \tag{2.24}$$

In this way, the brachistochrone problem can be formulated from a variational point of view as:

Let V be the space of all smooth functions $y(x)$ defined on an interval $[x_a, x_b]$, which are subject to the boundary conditions $y(x_a) = y_a$ and $y(x_b) = y_b$. Find a function $y^*(x) \in V$ for which the functional

$$T[y(x)] = \frac{1}{\sqrt{2g}} \int_{x_a}^{x_b} \sqrt{\frac{1 + [y'(x)]^2}{y_a - y(x)}}dx, \tag{2.25}$$

defined on V, takes on a minimum value.

The functional to be varied is thus

$$\widehat{T}[y(x), y'(x)] = \sqrt{\frac{1 + y'(x)^2}{y_a - y(x)}}dx. \tag{2.26}$$

Making use of the Euler-Lagrange equation, the set of parametric equations $(x, y)^*(\theta)$ for the brachistochrone are given by

$$x^*(\theta) = x_a + r(\theta - \sin(\theta)), \tag{2.27}$$

$$y^*(\theta) = y_a - r(1 - \cos(\theta)), \tag{2.28}$$

for $\theta \in [\theta_a, \theta_b]$. These are the equations of a cycloid, which is the locus of a point on the rim of a circle rolling along a horizontal line. By adjustments of the constants θ_a, θ_b and r, it is possible to construct one and only one cycloid which passes trough the points (x_a, y_a) and (x_b, y_b).

Equations (2.27) and (2.28) provide a descent time which is an absolute minimum when compared with all other arcs. This descent time can be obtained as

$$\begin{aligned} T[(x, y)^*(\theta)] &= \sqrt{\frac{1}{2g}} \int_0^{\theta_b} \sqrt{\frac{1 + \frac{\sin^2(\theta)}{(1 - \cos(\theta))^2}}{r(1 - \cos(\theta))}} r(1 - \cos(\theta))d\theta \\ &= \sqrt{\frac{r}{g}}\theta_b. \end{aligned} \tag{2.29}$$

Taking, for example, the points A and B to be $A = (0, 1)$ and $B = (1, 0)$, Equations (2.27) and (2.28) become

$$x^*(\theta) = 0.583(\theta - \sin(\theta)), \tag{2.30}$$
$$y^*(\theta) = 1 - 0.583(1 - \cos(\theta)), \tag{2.31}$$

for $\theta \in [0, 2.412]$. Figure 2.4 shows the shape of the brachistochrone for this particular example.

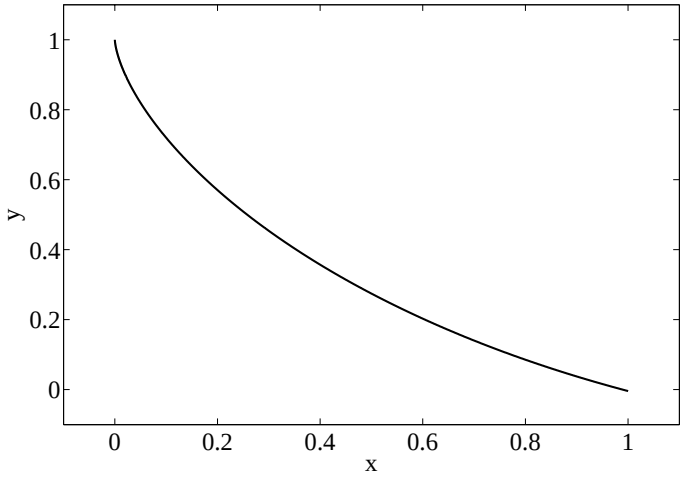


Figure 2.4: Analytical solution to the brachistochrone problem.

Finally, Equation (2.29) gives the descent time for this chute

$$T[(x, y)^*(\theta)] = 0.577. \tag{2.32}$$

2.3 The simplest constrained variational problem

In the simplest variational problem we have considered so far, the class of admissible curves was specified, apart from smoothness requirements, by boundary conditions. However, there are many applications of the calculus of variations in which constraints are imposed on the admissible curves. Such constraints are expressed as functionals.

Problem 2 (The simplest constrained variational problem)**.** Let V be the space of all smooth functions $y(x)$ defined on an interval $[x_a, x_b]$ which satisfy the boundary conditions $y(x_a) = y_a$ and $y(x_b) = y_b$, and such

$$C[y(x)] = \int_{x_a}^{x_b} \widehat{C}[x, y(x), y'(x)]dx = 0. \tag{2.33}$$

Find a function $y^*(x) \in V$ for which the functional defined on V

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx \tag{2.34}$$

takes on a minimum or a maximum value.

In other words, the simplest constrained variational problem consists of finding an extremum of a functional of the form (2.34), where the class of admissible functions consists of all smooth curves joining two points and satisfying some constraint of the form (2.33).

The most common procedure for solving the simplest constrained variational problem is the use of Lagrange multipliers [36]. In this method, the constrained problem is converted to an unconstrained problem. The resulting unconstrained problem can then be solved using the Euler-Lagrange equation.

Theorem 3 (Lagrange multipliers)**.** Let V be the space of all smooth functions $y(x)$ defined on an interval $[x_a, x_b]$ which satisfy the boundary conditions $y(a) = y_a$, $y(b) = y_b$, and such

$$\begin{aligned} C[y(x)] &= \int_{x_a}^{x_b} \widehat{C}[x, y(x), y'(x)]dx \\ &= 0. \end{aligned} \tag{2.35}$$

Let $y^*(x)$ be an extremal for the functional defined on V

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx. \tag{2.36}$$

Then, there exists a constant λ such that $y^*(x)$ is also an extremal of the functional

$$\begin{aligned} \bar{F}[y(x)] &= \int_{x_a}^{x_b} \left(\widehat{F}[x, y(x), y'(x)] + \lambda \widehat{C}[x, y(x), y'(x)]\right) dx \\ &= \int_{x_a}^{x_b} \widehat{\bar{F}}[x, y(x), y'(x)]dx. \end{aligned} \tag{2.37}$$

The constant λ is called the Lagrange multiplier.

The simplest constrained variational problem can then be solved by means of the Euler-Lagrange equation for the functional $\bar{F}$,

$$\frac{\partial \widehat{\bar{F}}}{\partial y} - \frac{d}{dx}\frac{\partial \widehat{\bar{F}}}{\partial y'} = 0. \tag{2.38}$$

Example 8 (The isoperimetric problem)**.** Dido's problem, which is also known as the isoperimetric problem, is be solved by means of Lagrange multipliers and the Euler-Lagrange equation:

Of all simple closed curves in the plane of a given length l, which encloses the maximum area?

Figure 8 states graphically this problem.

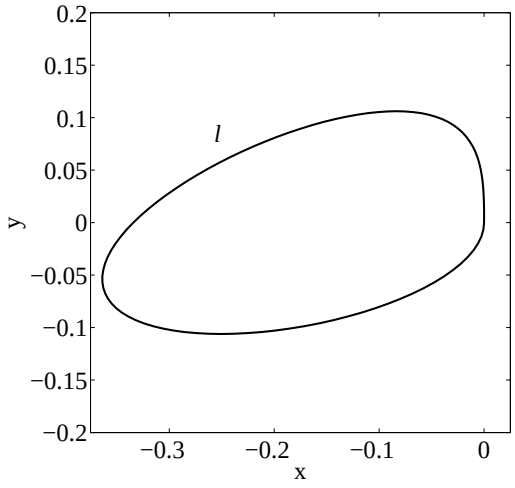


Figure 2.5: The isoperimetric problem statement.

Here we can not use a function $y(x)$ to specify the curve since closed curves will necessarily make the function multi-valued. Instead, we use the parametric equations $x = x(t)$ and $y = y(t)$, for $t \in [0, 1]$, and such that $x(0) = x(1)$, $y(0) = y(1)$ (where no further intersections occur).

For a plane curve specified in parametric equations as $(x, y)(t)$, the arc length is given by

$$l = \int_0^1 \sqrt{[x'(t)]^2 + [y'(t)]^2}dt. \tag{2.39}$$

On the other hand, Green's theorem gives the signed area as

$$a = \frac{1}{2}\int_0^1 (x(t)y'(t) - x'(t)y(t))\, dt. \tag{2.40}$$

Thus, the isoperimetric problem can be formulated from a variational point of view as follows:

Let V be the space of all smooth functions $(x, y)(t)$ defined on the interval $[0, 1]$, which satisfy the boundary conditions $x(0) = 0$, $x(1) = 0$, $y(0) = 0$, $y(1) = 0$, and such

$$\begin{aligned} L[(x, y)(t)] &= \int_0^1 \sqrt{[x'(t)]^2 + [y'(t)]^2}dt - l \\ &= 0. \end{aligned} \tag{2.41}$$

Find a function $(x, y)^*(t) \in V$ for which the functional defined on V

$$A[(x, y)(t)] = \frac{1}{2}\int_0^1 (x(t)y'(t) - x'(t)y(t))\, dt \tag{2.42}$$

takes on a maximum value.

We can reformulate this constrained variational problem as an unconstrained variational problem by means of Lagrange multipliers:

Let V be the space of all smooth functions $(x, y)(t)$ defined on an interval $[0, 1]$, which satisfy the boundary conditions $x(0) = 0$, $x(1) = 0$, $y(0) = 0$, $y(1) = 0$. Find a function $(x, y)^*(t) \in V$ for which the functional defined on V

$$F[(x, y)(t)] = \int_0^1 \left(\sqrt{[x'(t)]^2 + [y'(t)]^2} + \lambda(x(t)y'(t) - x'(t)y(t)) \right) dt \tag{2.43}$$

takes on a maximum value.

The functional to be varied is therefore

$$\begin{aligned} \widehat{F}[x(t), x'(t), y(t), y'(t)] &= \sqrt{[x'(t)]^2 + [y'(t)]^2} \\ &+ \lambda(x(t)y'(t) - x'(t)y(t)). \end{aligned} \tag{2.44}$$

Integrating the Euler-Lagrange equation here a circle with radius r is obtained. A set of parametric equations $(x, y)^*(t)$ for the circle is given by

$$x^*(t) = x_a + r\cos(2\pi t), \tag{2.45}$$

$$y^*(t) = x_b + r\sin(2\pi t), \tag{2.46}$$

for $t \in [0, 1]$. By adjustments of the constants a, b and r it is possible to construct a circle with perimeter l which satisfies the boundary conditions $(x(0), y(0)) = (0, 0)$ and $(x(1), y(1)) = (0, 0)$.

Taking, for instance, the perimeter of the circle to be $l = 1$, Equations (2.45) and (2.46) become

$$x^*(t) = -\frac{1}{2\pi} + \frac{1}{2\pi}\cos(2\pi t), \tag{2.47}$$

$$y^*(t) = \frac{1}{2\pi}\sin(2\pi t), \tag{2.48}$$

for $t \in [0, 1]$. Figure 8 shows the shape of the circle for this particular example.

The area of such a circle is

$$A[(x, y)^*(t)] = \frac{1}{4\pi}. \tag{2.49}$$

2.4 Direct methods in variational problems

While many variational problems can be solved analytically by means of the Euler-Lagrange equation, the only practical technique for general variational problems is to approximate the solution using 'direct methods' [29].

The fundamental idea underlying the so called direct methods is to consider a variational problem as a limit problem for some function optimization problem in many dimensions. This new problem is then solved by usual methods. In this way, the difference between variational problems and function optimization problems is in the number of dimensions. Variational problems entail infinite dimensions, while function optimization problems involve finite dimensions.

In this section we describe two direct methods, the Euler method and the Ritz method.

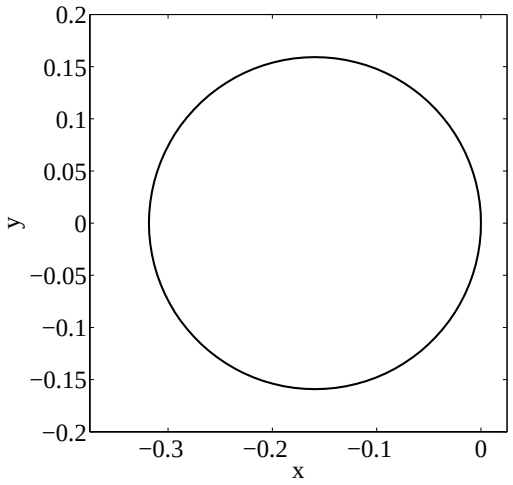


Figure 2.6: Analytical solution to the isoperimetric problem.

The Euler method

The first direct method studied here is the Euler method, also called the finite differences method [29]. In order to understand the Euler method, consider the simplest variational problem:

Let V be the space of all smooth functions $y(x)$ defined on an interval $[a, b]$ which satisfy the boundary conditions $y(x_a) = y_a$, $y(x_b) = y_b$. Find a function $y^*(x) \in V$ for which the functional defined on V

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx \tag{2.50}$$

takes on a minimum or a maximum value.

The main idea of the Euler method is that the values of the functional $F[y(x)]$ are considered not along all admissible curves $y(x) \in V$, but along polygonal curves which consist of a number $n - 1$ of line segments with vertices at points

$$(x_a, y_a), (x_1, y_1), \ldots, (x_n, y_n), (x_b, y_b)$$

where $x_i = x_a + ih$, for $i = 1, \ldots, n$ and being $h = (x_b - x_a)/(n + 1)$, see Figure 2.7.

Along such polygonal curves the functional $F[y(x)]$ turns into a function $f(y_1, \ldots, y_n)$. The problem is then to choose $y_1, \ldots, y_n$ so that the function $f(y_1, \ldots, y_n)$ has an extremum. The necessary condition for the ordinates $y_1, \ldots, y_n$ to be an extremal of f is that they satisfy

$$\begin{aligned} \nabla f(y_1, \ldots, y_n) &= \left(\frac{\partial f}{\partial y_1}, \ldots, \frac{\partial f}{\partial y_n} \right) \\ &= 0. \end{aligned} \tag{2.51}$$

By doing so we shall have a polygonal curve which is an approximate solution to the variational problem in question.

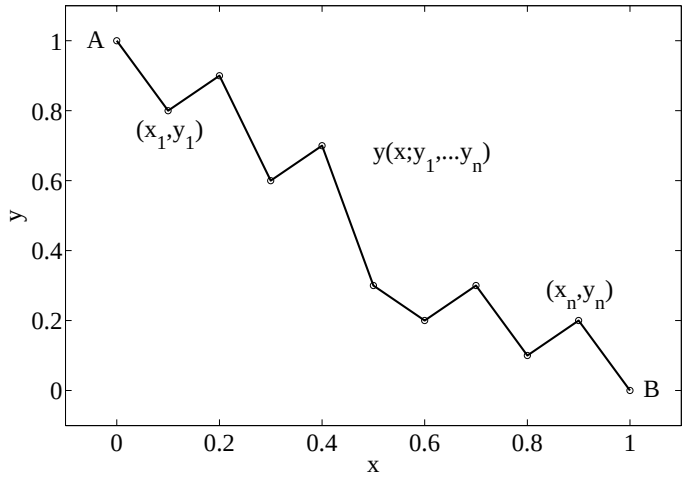


Figure 2.7: Illustration of the Euler method.

The Ritz method

The second direct method studied in this section is the Ritz method [29].As before, in order to understand the Ritz method, consider the simplest variational problem:

Let V be the space of all smooth functions $y(x)$ defined on an interval $[a, b]$ which satisfy the boundary conditions $y(x_a) = y_a$, $y(x_b) = y_b$. Find a function $y^*(x) \in V$ for which the functional defined on V

$$F[y(x)] = \int_{x_a}^{x_b} \widehat{F}[x, y(x), y'(x)]dx \tag{2.52}$$

takes on a minimum or a maximum value.

The idea of the Ritz method is that the values of the functional $F[y(x)]$ are considered not along all admissible curves $y(x) \in V$, but only along all possible linear combinations of a certain sequence of n functions $g_1(x), \ldots, g_n(x)$,

$$y(x) = \sum_{i=1}^{n} \alpha_i g_i(x). \tag{2.53}$$

The elements in the function space defined by Equation (2.53) must satisfy the boundary conditions for a given problem, which is a restriction on the choice of the sequence of functions $g_i(x)$. Along such linear combination the functional $F[y(x)]$ becomes a function of the coefficients $f(\alpha_1, \ldots, \alpha_n)$. The problem is the to choose $\alpha_1, \ldots, \alpha_n$ so that the function $f(\alpha_1, \ldots, \alpha_n)$ has an extremum. The necessary condition for the coefficients $\alpha_1, \ldots, \alpha_n$ to be an extremal of f is that they satisfy

$$\begin{aligned} \nabla f(\alpha_1, \ldots, \alpha_n) &= \left(\frac{\partial f}{\partial \alpha_1}, \ldots, \frac{\partial f}{\partial \alpha_n} \right) \\ &= 0. \end{aligned} \tag{2.54}$$

By doing so we shall obtain a curve which is an approximate solution to the variational problem in question. The initial choice of the functions $\{g_1, \ldots, g_n\}$, which are called *coordinate functions* is of greet importance, and therefore a successful application of Ritz's method depends on an adequate choice of the coordinate functions. Figure 2.8 illustrates this method.

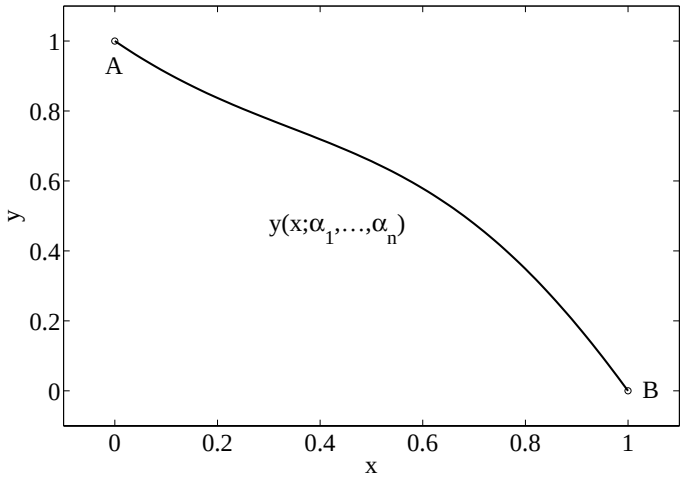


Figure 2.8: Illustration of the Ritz method.

This page is intentionally left blank.

Chapter 3

A variational formulation for the multilayer perceptron

There are many different types of neural networks. The multilayer perceptron is an important one, and most of the literature in the field is referred to this neural network. In this Chapter, the learning problem in the multilayer perceptron is formulated from a variational point of view. In this way, learning tasks lie in terms of finding a function which causes some functional to assume an extreme value. As we shall see, the multilayer perceptron provides a general framework for solving variational problems.

3.1 Introduction

The multilayer perceptron is characterized by a neuron model, a network architecture and associated objective functionals and training algorithms. This four concepts are briefly described next.

1. *Neuron model*. A neuron model is a mathematical model of the behavior of a single neuron in a biological nervous system. The characteristic neuron model in the multilayer perceptron is the so called perceptron. The perceptron neuron model receives information in the form of a set of numerical input signals. This information is then integrated with a set of free parameters to produce a message in the form of a single numerical output signal.

2. *Network architecture*. In the same way a biological nervous system is composed of interconnected biological neurons, an artificial neural network is built up by organizing artificial neurons in a network architecture. In this way, the architecture of a network refers to the number of neurons, their arrangement and connectivity. The characteristic network architecture in the multilayer perceptron is the so called feed-forward architecture.

3. *Objective functional*. The objective functional plays an important role in the use of a neural network. It defines the task the neural network is required to do and provides a measure of the quality of the representation that the network is required to learn. The choice of a suitable objective functional depends on the particular application.

4. *Training algorithm*. The procedure used to carry out the learning process is called training algorithm, or learning algorithm. The training algorithm is applied to the network to in

order to obtain a desired performance. The type of training is determined by the way in which the adjustment of the free parameters in the neural network takes place.

Figure 3.1 depicts an activity diagram for the learning problem in the multilayer perceptron. The solving approach here consists of three steps. The first step is to choose a suitable parameterized function space in which the solution to the problem is to be approximated. The elements of this family of functions are those spanned by a multilayer perceptron. In the second step the variational problem is formulated by selecting an appropriate objective functional, defined on the function space chosen before. The third step is to solve the reduced function optimization problem. This is performed with a training algorithm capable of finding an optimal set of parameters.

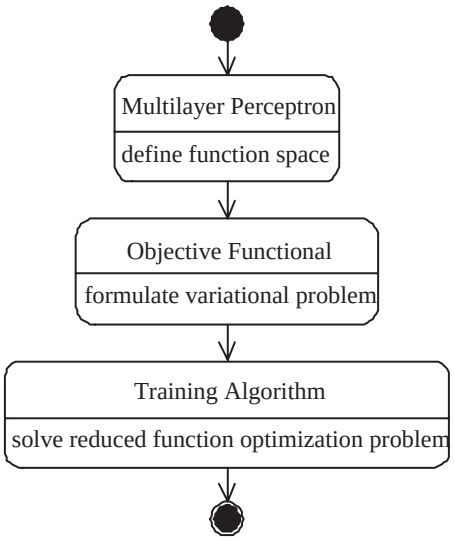


Figure 3.1: Activity diagram for the learning problem in the multilayer perceptron.

3.2 The perceptron neuron model

As we have said, a neuron model is the basic information processing unit in a neural network. The perceptron is the characteristic neuron model in the multilayer perceptron. Following current practice [98], the term perceptron is here applied in a more general way than by Rosenblatt, and covers the types of units that were later derived from the original perceptron.

3.2.1 The elements of the perceptron

The block diagram in Figure 3.2 is a graphical representation of the perceptron. Here we identify three basic elements, which transform the input signals $(x_1, \ldots, x_n)$ in a single output signal y [13]:

- A set of free parameters $\underline{\alpha}$, which consists of a bias b and a vector of synaptic weights $(w_1, \ldots, w_n)$.
- A combination function h, which combines the input signals and the free parameters to produce a single net input signal u.

- An activation function or transfer function g, which takes as argument the net input signal and produces the output signal.

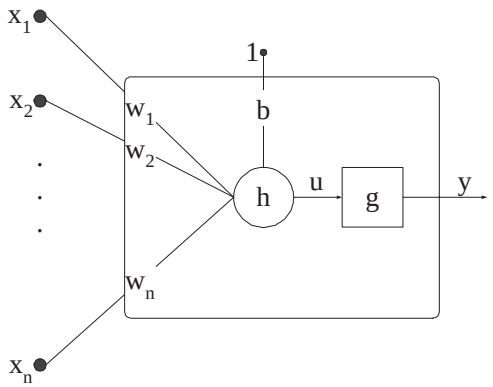


Figure 3.2: Perceptron neuron model.

Next we describe in more detail the three basic elements of this neuron model [41]:

Free parameters

The free parameters allow a neuron model to be trained to perform a task. In the perceptron, the set of free parameters is:

$$\underline{\alpha} = (b, \mathbf{w}) \in \mathbb{R} \times \mathbb{R}^n \tag{3.1}$$

where b is called the bias and $\mathbf{w} = (w_1, \ldots, w_n)$ is called the synaptic weight vector. Note then that the number of free parameters of this neuron model is $1 + n$, where n is the number of inputs in the neuron.

Combination function

In the perceptron, the combination function h computes the inner product of the input vector $\mathbf{x} = (x_1, \ldots, x_n)$ and the synaptic weight vector $\mathbf{w} = (w_1, \ldots, w_n)$ to produce a net input signal u. This model also includes a bias externally applied, denoted by b, which increases or reduces the net input signal to the activation function, depending on whether it is positive or negative, respectively. The bias is often represented as a synaptic weight connected to an input fixed to $+1$,

$$h(\mathbf{x}; b, \mathbf{w}) \quad = \quad b + \sum_{i=1}^{n} w_i x_i. \tag{3.2}$$

Activation function

The activation function g defines the output signal y from the neuron in terms of its net input signal u. In practice we can consider many useful activation functions [27]. Three of the most used activation functions are the threshold function, the linear function and the sigmoid function [41].

Threshold function . For this activation function, represented in Figure 3.3 we have

$$g(u) = \begin{cases} -1 & u < 0 \\ 1 & u \geq 0 \end{cases} \tag{3.3}$$

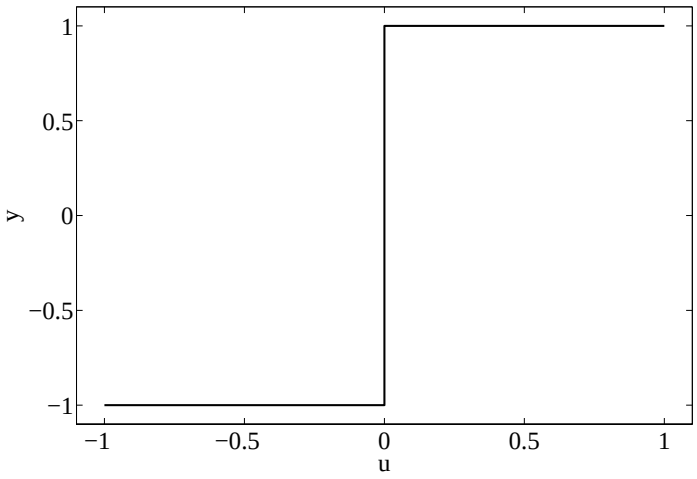


Figure 3.3: A threshold activation function.

That is, the threshold activation function limits the output of the neuron to -1 if the net input u is less than 0; or 1 if u is equal to or greater than 0.

There might be some cases when we need to compute the activation derivative of the neuron,

$$g'(u) \equiv \frac{dg(u)}{du}. \tag{3.4}$$

or its second derivative,

$$g''(u) \equiv \frac{d^2g(u)}{du^2}. \tag{3.5}$$

The problem with the threshold activation function is that it is not differentialbe at the point $u = 0$.

Sigmoid function . The sigmoid function is the most used activation function when constructing neural networks. Is an monotonous crescent function which exhibits a good balance between a linear and a non-linear behavior. An example of a sigmoid function is the hyperbolic tangent function, defined by

$$g(u) = \tanh(u). \tag{3.6}$$

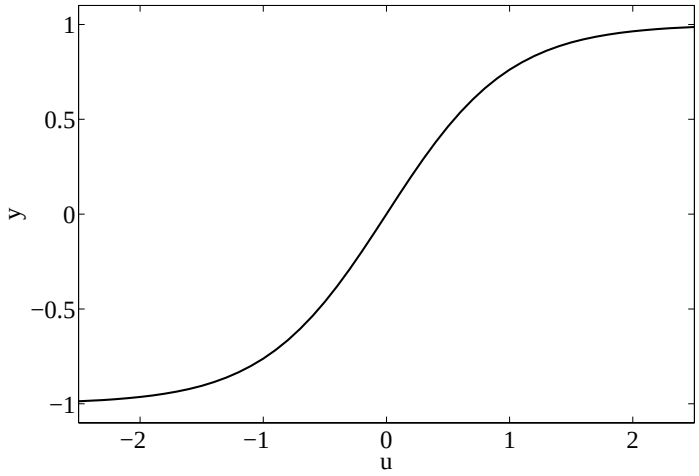


Figure 3.4: A sigmoid activation function.

The hyperbolic tangent function is represented in Figure 3.4.

For this sigmoid function, the activation derivative is given by

$$g'(u) = 1 - \tanh^2(u), \tag{3.7}$$

and the second derivative of the activation is given by

$$g''(u) = -2\tanh(u)(1 - \tanh^2(u)). \tag{3.8}$$

Linear function . For the linear activation function, described in Figure 3.5 we have

$$g(u) = u. \tag{3.9}$$

Thus, the output signal of a neuron model with linear activation function is equal to its net input.

For the linear function, the activation derivative is given by

$$g'(u) = 1, \tag{3.10}$$

and the second derivative of the activation is given by

$$g''(u) = 0. \tag{3.11}$$

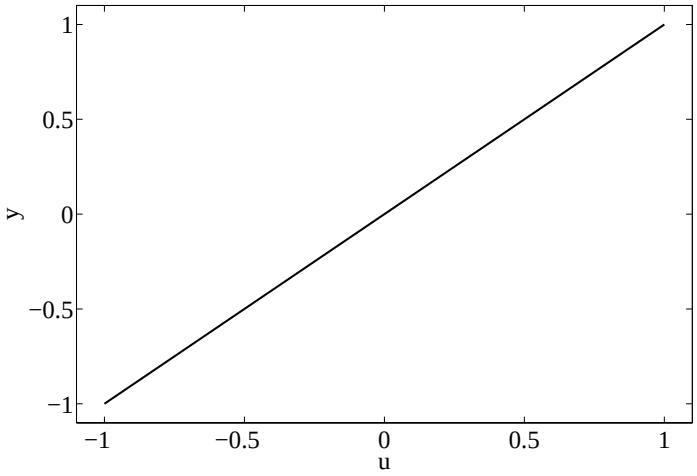


Figure 3.5: A linear activation function.

3.2.2 The perceptron function space

Mathematically, a perceptron neuron model may be viewed as a parameterized function space V from an input $X \subset \mathbb{R}^n$ to an output $Y \subset \mathbb{R}$. Elements of V are parameterized by the bias and the vector of synaptic weights of the neuron. In this way the dimension of V is $n + 1$.

We can write down an expression for the elements of the function space which a perceptron can define [15]. The net input to the neuron is obtained by first forming a linear combination of the input signals and the synaptic weights, and then adding the bias, to give

$$u = b + \sum_{i=1}^{n} w_i x_i. \tag{3.12}$$

The output of the neuron is obtained transforming the linear combination in Equation (3.12) with an activation function g to give

$$y(\mathbf{x}; b, \mathbf{w}) = g\left(b + \sum_{i=1}^{n} w_i x_i\right). \tag{3.13}$$

Distinct activation functions cause distinct families of functions which a perceptron can define. Similarly, distinct sets of free parameters cause distinct elements in the function space which a specific perceptron defines. The concepts of linear and sigmoid perceptron are implemented within the classes LinearPerceptron and SigmoidPerceptron of the C++ library Flood [61].

Although a single perceptron can solve some simple learning tasks, the power of neural computation comes from connecting many neurons in a network architecture [98].

3.3 The multilayer perceptron network architecture

Neurons can be combined to form a neural network. The architecture of a neural network refers to the number of neurons, their arrangement and connectivity. Any network architecture can be

symbolized as a directed and labeled graph, where nodes represent neurons and edges represent connectivities among neurons. An edge label represents the free parameter of the neuron for which the flow goes in [13].

Most neural networks, even biological neural networks, exhibit a layered structure. In this work layers are the basis to determine the architecture of a neural network [98]. Thus, a neural network typically consists on a set of sensorial nodes which constitute the input layer, one or more hidden layers of neurons and a set of neurons which constitute the output layer.

As it was said above, the characteristic neuron model of the multilayer perceptron is the perceptron. On the other hand, the multilayer perceptron has a feed-forward network architecture.

3.3.1 Feed-forward architectures

Feed-forward architectures contain no cycles, i.e., the architecture of a feed-forward neural network can then be represented as an acyclic graph. Hence, neurons in a feed-forward neural network are grouped into a sequence of $c+1$ layers $L^{(1)}, \ldots, L^{(c+1)}$, so that neurons in any layer are connected only to neurons in the next layer. The input layer $L^{(0)}$ consists of n external inputs and is not counted as a layer of neurons; the hidden layers $L^{(1)}, \ldots, L^{(c)}$ contain $h_1, \ldots, h_c$ hidden neurons, respectively; and the output layer $L^{(c+1)}$ is composed of m output neurons. Communication proceeds layer by layer from the input layer via the hidden layers up to the output layer. The states of the output neurons represent the result of the computation [98].

Figure 3.6 shows a feed-forward architecture, with n inputs, c hidden layers with h_i neurons, for $i = 1, \ldots, c$, and m neurons in the output layer.

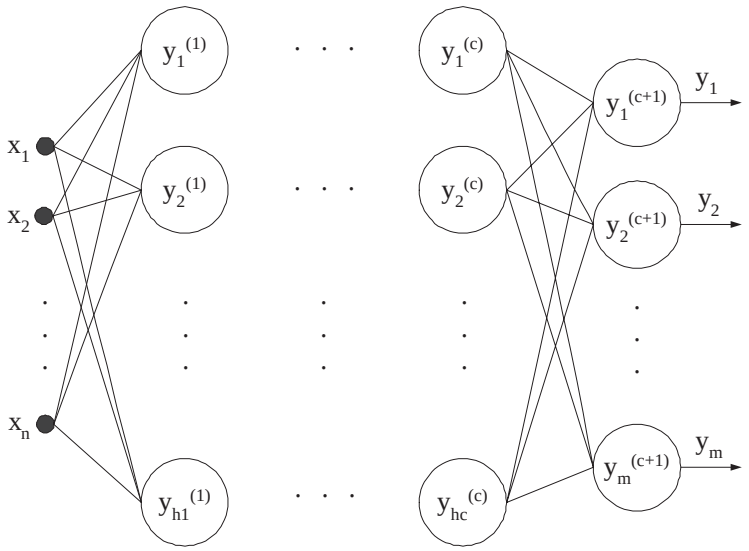


Figure 3.6: The feed-forward network architecture.

In this way, in a feed-forward neural network, the output of each neuron is a function of the inputs. Thus, given an input to such a neural network, the activations of all neurons in the output layer can be computed in a deterministic pass [15].

3.3.2 The multilayer perceptron function space

In Section 3.2 we considered the space of functions that a perceptron neuron model can define. As it happens with a single perceptron, a multilayer perceptron neural network may be viewed as a parameterized function space V from an input $X \subset \mathbb{R}^n$ to an output $Y \subset \mathbb{R}^m$. Elements of V are parameterized by the biasses and synaptic weights in the neural network, which can be grouped together in a s-dimensional vector $\underline{\alpha} = (\alpha_1, \ldots, \alpha_s)$. The dimension of the function space V is therefore s.

Figure 3.7 shows a multilayer perceptron, with n inputs, one hidden layer with h_1 neurons and m neurons in the output layer. Biases in the hidden layer are represented as synaptic weights from an extra input with a fixed value of $x_0 = 1$. Similarly, biases in the output layer are represented as synaptic weights from an extra hidden neuron, with a fixed activation also fixed to $y_0^{(1)} = 1$.

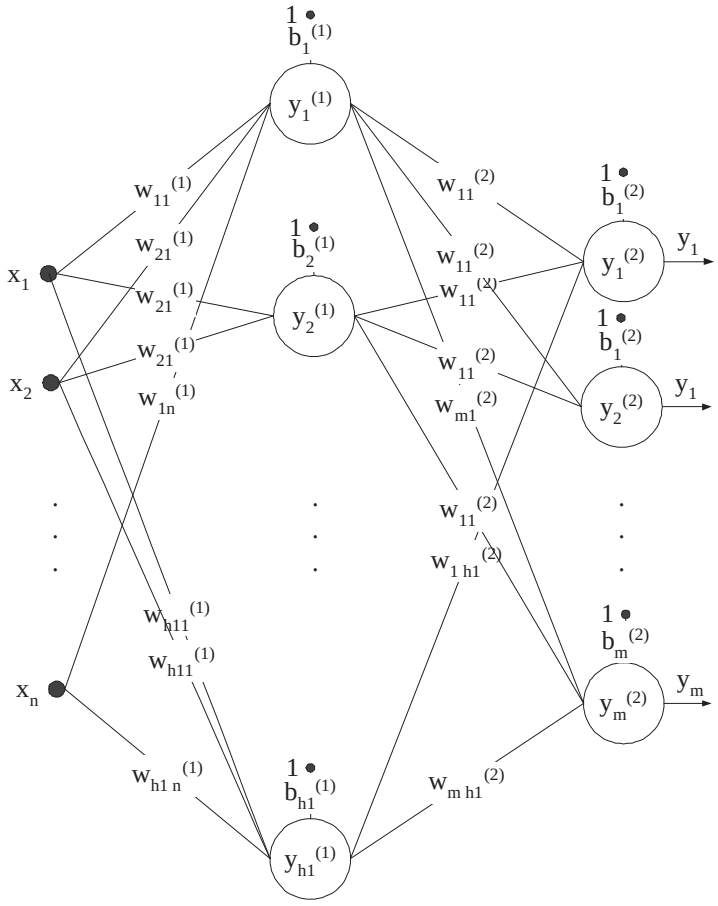


Figure 3.7: A two-layer perceptron.

We can write down the analytical expression for the elements of the function space which the multilayer perceptron shown in Figure 3.7 can define [15]. Indeed, the net input to the hidden neuron j is obtained first by forming a linear combination of the n input signals, and adding the bias, to give

$$\begin{aligned} u_j^{(1)} &= b_j^{(1)} + \sum_{i=1}^{n} w_{ji}^{(1)} x_i \\ &= \sum_{i=0}^{n} \alpha_{ji}^{(1)} x_i, \end{aligned} \tag{3.14}$$

for $j = 1, \ldots, h_1$. $\alpha_{j0}^{(1)}$ denotes the bias of neuron j in layer (1), for which $x_0 = 1$; $\alpha_{ji}^{(1)}$, for $i = 1, \ldots, n$, denotes the synaptic weight of neuron j in layer (1) which comes from input i.

The output of hidden neuron j is obtained transforming the linear combination in Equation (3.14) with an activation function $g^{(1)}$ to give

$$y_j^{(1)} = g^{(1)}\left(u_j^{(1)}\right), \tag{3.15}$$

for $j = 1, \ldots, h_1$. The neural network outputs are obtained transforming the output signals of the neurons in the hidden layer by the neurons in the output layer. Thus, the net input for each output neuron k is obtained forming a linear combination of the output signals from the hidden layer neurons of the form

$$\begin{aligned} u_k^{(2)} &= b_k^{(2)} + \sum_{j=1}^{h_1} w_{kj}^{(2)} y_j^{(1)} \\ &= \sum_{j=0}^{h_1} \alpha_{kj}^{(2)} y_j^{(1)}, \end{aligned} \tag{3.16}$$

for $k = 1, \ldots, m$. The value $\alpha_{k0}^{(2)}$ denotes the bias of neuron k in layer (2), for which $y_0^{(1)} = 1$; similarly, the value $\alpha_{kj}^{(2)}$, $j = 1, \ldots, h_1$, denotes the synaptic weight of neuron k in layer (2) which comes from input j.

The activation of the output neuron k is obtained transforming the linear combination in Equation (3.16) with an activation function $g^{(2)}$ to give

$$y_k = g^{(2)}\left(u_k^{(2)}\right), \tag{3.17}$$

for $k = 1, \ldots, m$. Note that the activation function in the output layer does not need to be the same than for that in the hidden layer.

If we combine (3.14), (3.15), (3.16) and (3.17) we obtain an explicit expression for the function represented by the neural network diagram in Figure 3.7 of the form

$$y_k = g^{(2)}\left(\sum_{j=0}^{h_1} \alpha_{kj}^{(2)} g^{(1)}\left(\sum_{i=0}^{n} \alpha_{ji}^{(1)} x_i\right)\right), \tag{3.18}$$

for $k = 1, \ldots, m$. In this way, the multilayer perceptron can be considered as a function of many variables composed by superposition and addition of functions of one variable. The neural network

shown in Figure 3.7 corresponds to a transformation of the input variables by two successive different networks with a single layer. This class of networks can be extended by considering new successive transformation of the same kind, which corresponds to networks with more hidden layers.

Distinct activation functions cause distinct families of functions which a multilayer perceptron can define. Similarly, distinct sets of free parameters cause distinct elements in the function space which a specific multilayer perceptron defines.

A multilayer perceptron with a sigmoid hidden layer and a linear output layer is implemented within the C++ class MultilayerPerceptron of Flood [61].

3.3.3 The Jacobian matrix

There are some cases when, in order to evaluate the objective functional of a multilayer perceptron, we need to compute the derivatives of the network outputs with respect to the inputs. That derivatives can be grouped together in the Jacobian matrix, whose elements are given by

$$J_{ki} \equiv \frac{\partial y_k}{\partial x_i}, \tag{3.19}$$

for $i = 1, \ldots, n$, $k = 1, \ldots, m$ and where each such derivative is evaluated with all other inputs held fixed.

In Sections 5.2 and 5.4 we make use of the Jacobian matrix to evaluate two different objective functionals for the multilayer perceptron.

The Jacobian matrix can be evaluated either by using a back-propagation procedure, or by means of numerical differentiation.

The back-propagation algorithm for the calculus of the Jacobian matrix

Here we evaluate the Jacobian matrix for the multilayer perceptron using the back-propagation algorithm. Results are derived for the one hidden layer perceptron shown in Figure 3.7, but they are easily generalized to networks with several hidden layers of neurons.

We start by writing the element J_{ki} in the form

$$\begin{aligned} J_{ki} &= \frac{\partial y_k}{\partial x_i} \\ &= \sum_{j=1}^{h_1} \frac{\partial y_k}{\partial u_j^{(1)}} \frac{\partial u_j^{(1)}}{\partial x_i}, \end{aligned} \tag{3.20}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, m$. In order to determine the derivative $\partial u_j^{(1)} / \partial x_i$ let us consider the net input signal for each neuron in the hidden layer

$$u_j^{(1)} = \sum_{i=0}^{n} \alpha_{ji}^{(1)} x_i, \tag{3.21}$$

for $j = 1, \ldots, h_1$. Thus, the mentioned derivative yields

$$\frac{\partial u_j^{(1)}}{\partial x_i} = \alpha_{ji}^{(1)}, \tag{3.22}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, h_1$. Equation (3.20) becomes

$$J_{ki} = \sum_{j=1}^{h_1} \alpha_{ij}^{(1)} \frac{\partial y_k}{\partial u_j^{(1)}}, \tag{3.23}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, m$. We now write down a recursive back-propagation formula to determine the derivatives $\partial y_k / \partial u_j^{(1)}$.

$$\frac{\partial y_k}{\partial u_j^{(1)}} = \sum_{l=1}^{m} \frac{\partial y_k}{\partial u_l^{(2)}} \frac{\partial u_l^{(2)}}{\partial u_j^{(1)}}, \tag{3.24}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. The derivative $\partial u_l^{(2)} / \partial u_j^{(1)}$ can be calculated by first considering the net input signal of each neuron in the output layer

$$u_l^{(2)} = \sum_{j=0}^{h_1} \alpha_{lj}^{(2)} y_j^{(1)}, \tag{3.25}$$

for $l = 1, \ldots, m$. The activation of each neuron in the hidden layer is given by

$$y_j^{(1)} = g^{(1)}(u_j^{(1)}), \tag{3.26}$$

for $j = 1, \ldots, h_1$. So we can write an expression for the net input signal of each neuron in the output layer in the form

$$u_l^{(2)} = \sum_{j=0}^{h_1} \alpha_{lj}^{(2)} g^{(1)}(u_j^{(1)}), \tag{3.27}$$

for $l = 1, \ldots, m$. Thus, the derivative $\partial u_l^{(2)} / \partial u_j^{(1)}$ yields

$$\frac{\partial u_l^{(2)}}{\partial u_j^{(1)}} = w_{lj}^{(2)} g'^{(1)}(u_j^{(1)}), \tag{3.28}$$

for $j = 1, \ldots, h_1$ and $l = 1, \ldots, m$. Equation (3.21) becomes

$$\frac{\partial y_k}{\partial u_j^{(1)}} = g'(u_j^{(1)}) \sum_{l=1}^{m} \alpha_{lj}^{(2)} \frac{\partial y_k^{(2)}}{\partial u_l^{(2)}}, \tag{3.29}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. For the output neurons we have

$$y_k = g^{(2)}(u_k^{(2)}), \tag{3.30}$$

for $k = 1, \ldots, m$, and from which

$$\frac{\partial y_k}{\partial u_{k'}^{(2)}} = g'^{(2)}(u_k^{(2)})\delta_{kk'}. \tag{3.31}$$

$\delta_{kk'}$ is the Kronecker delta symbol, and equals 1 if $k = k'$ and 0 otherwise. We can therefore summarize the procedure for evaluating the Jacobian matrix as follows [15]:

1. Apply the input vector corresponding to the point in input space at which the Jacobian matrix is to be found, and forward propagate in the usual way to obtain the activations of all of the hidden and output neurons in the network.

2. For each row k of the Jacobian matrix, corresponding to the output neuron k, back-propagate to the hidden units in the network using the recursive relation (3.29), and starting with (3.31).

3. Use (3.23) to do the back-propagation to the inputs. The second and third steps are the repeated for each value of k, corresponding to each row of the Jacobian matrix.

A C++ software implementation of this algorithm can be found inside the class MultilayerPerceptron of Flood [61].

Numerical differentiation for the calculus of the Jacobian matrix

The Jacobian matrix for the multilayer perceptron J_{ki} can also be evaluated using numerical differentiation [15]. This can be done by perturbing each input in turn, and approximating the derivatives by using forward differences,

$$\frac{\partial y_k}{\partial x_i} = \frac{y_k(x_i + \epsilon) - y_k(x_i)}{\epsilon} + \mathcal{O}(\epsilon), \tag{3.32}$$

for $i = 1, \ldots, n$, $k = 1, \ldots, m$ and for some small numerical value of ϵ.

The accuracy of the finite differences method can be improved significantly by using central differences of the form

$$\frac{\partial y_k}{\partial x_i} = \frac{y_k(x_i + \epsilon) - y_k(x_i - \epsilon)}{2\epsilon} + \mathcal{O}(\epsilon^2), \tag{3.33}$$

also for $i = 1, \ldots, n$, $k = 1, \ldots, m$ and for some small numerical value of ϵ.

In a software implementation the Jacobian matrix for the multilayer perceptron J_{ki} should be evaluated using the back-propagation algorithm, since this gives the greatest accuracy and numerical efficiency [15]. The implementation of such algorithm should be checked by using numerical differentiation. Both forward and central differences for the Jacobian are also implemented inside the class MultilayerPerceptron of Flood [61].

3.3.4 Universal approximation

A multilayer perceptron with as few as one hidden layer of sigmoid neurons and an output layer of linear neurons provides a general framework for approximating any function from one finite dimensional space to another up to any desired degree of accuracy, provided sufficiently many hidden neurons are available. In this sense, multilayer perceptron networks are a class of universal approximators [49].

A detailed statement of the universal approximation theorem for the multilayer perceptron is out of the scope of this work, so that it is not included here. The interested reader can find this theorem, as well as its demonstration, in [49].

The universal approximation capability of the multilayer perceptron implies that any lack of success in an application must arise from a wrong number of hidden neurons, the lack of the objective functional or inadequate training.

3.3.5 Pre and post-processing

In practice it is always convenient to pre-process the inputs in order to produce input signals to the neural network of order zero. In this way, if all the free parameters are of order zero, the output signals will be also of order zero.

There are several pre and post-processing methods. Two of the most used are the mean and standard deviation and the minimum and maximum pre and post-processing method. Note that these two forms of pre and post-processing will affect the back-propagation algorithm for the Jacobian matrix.

The mean and standard deviation pre and post-processing method

With this method the inputs are pre-processed so that the input signals have mean 0 and standard deviation 1,

$$\hat{x}_i = \frac{x_i - \rho(x_i)}{\sigma(x_i)}, \tag{3.34}$$

for $i = 1, \ldots, n$, and where x_i are the inputs, $\hat{x}_i$ are the input signals to the neural network, and $\rho(x_i)$ and $\sigma(x_i)$ are an estimation of the mean and the standard deviation of the input variables, respectively.

The output signals from the neural network are then post-processed to produce the outputs

$$y_i = \rho(y_i) + \sigma(y_i)\hat{y}_i, \tag{3.35}$$

for $i = 1, \ldots, m$, and where y_i are the outputs, $\hat{y}_i$ are the output signals from the network, and $\rho(y_i)$ and $\sigma(y_i)$ are an estimation of the mean and the standard deviation of the output variables, respectively.

The minimum and maximum pre and post-processing method

Here the inputs are pre-processed to produce input signals whose minimum and maximum values are -1 and 1,

$$\hat{x}_i = 2\frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} - 1, \tag{3.36}$$

for $i = 1, \ldots, n$, and where x_i are the inputs, $\hat{x}_i$ are the input signals to the network, and $\min(x_i)$ and $\max(x_i)$ are an estimation of the minimum and the maximum values of the input variables, respectively.

The output signals from the network are then post-processed to produce the outputs

$$y_i = 0.5\,(\hat{y}_i + 1)\,(\max(y_i) - \min(y_i)) + \min(y_i), \tag{3.37}$$

for $i = 1, \ldots, m$, and where y_i are the outputs, $\hat{y}_i$ are the output signals from the network, and $\min(y_i)$ and $\max(y_i)$ are an estimation of the minimum and the maximum values of the outputs from the neural network, respectively.

3.3.6 Multilayer perceptron extensions

This Section describes some possible extensions for the multilayer perceptron. They include independent parameters, boundary conditions and lower and upper bounds. In some situations these extensions can improve the performance by the numerical method. In some other cases they allow to deal with applications which would be untractable otherwise. In summary, this augmented class of neural network might be able to span a more suited function space for some variational problems.

Independent parameters

If some information not related to input-output relationships is needed, then the problem is said to have independent parameters. They are not a part of the neural network, but they are associated to it. The independent parameters can be grouped together in a q-dimensional vector $\underline{\beta} = (\beta_1, \ldots, \beta_q)$.

Consequently, a multilayer perceptron with associated independent parameters spans a space of functions V from an input $X \subseteq \mathbb{R}^n$ to an output $Y \subseteq \mathbb{R}^m$, where n and m are the number of inputs and outputs, respectively. The elements of this family of functions are parameterized by both, the biases and synaptic weights vector, $\underline{\alpha} = (\alpha_1, \ldots, \alpha_p)$, and the independent parameters vector, $\underline{\beta} = (\beta_1, \ldots, \beta_q)$. The dimension of V is thus $p + q$, and the functions here are of the form

$$\begin{aligned} \mathbf{y} : X &\rightarrow Y \\ \mathbf{x} &\mapsto \mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}). \end{aligned}$$

Note that the total set of free parameters is composed by the biases and synaptic weights and the independent parameters. The first group defines the output from the neural network for a given input. The second group provides some separate sort of information.

Examples of problems in which independent parameters are used can be found in Sections 7.3 and 6.2.1 .

Boundary conditions

If some outputs are specified for given inputs, then the problem is said to include boundary conditions. A boundary condition between some input $x = a$ and some output $y = y_a$ is written

$y(a) = y_a$. In order to deal with boundary conditions the output signals from the neural network can be post-processed as follows

$$\mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}) = \varphi_0(\mathbf{x}) + \varphi_1(\mathbf{x})\mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}), \tag{3.38}$$

where the function $\varphi_0(\mathbf{x})$ is called a particular solution term and the function $\varphi_1(\mathbf{x})$ is called an homogeneous solution term. The first must hold $\varphi_0(a) = y_a$ if there is a condition $y(a) = y_a$. The second must hold $\varphi_1(a) = 0$ if there is a condition $y(a) = y_a$. It is easy to see that this approach makes all the elements of the function space to satisfy the boundary conditions. It is important to remark that the expressions of the particular and homogeneous solution terms depend on the problem at hand.

For the simplest case of one input and one output variables and one boundary condition $y(a) = y_a$, a possible set of particular and homogeneous solution terms could be

$$\varphi_0(x) = a, \tag{3.39}$$
$$\varphi_1(x) = x - a. \tag{3.40}$$

In the situation of one input and one output variables and two conditions $y(a) = y_a$ and $y(b) = y_b$ we could have

$$\varphi_0(x) = y_a + \frac{y_b - y_a}{b - a}x, \tag{3.41}$$
$$\varphi_1(x) = (x - a)(x - b). \tag{3.42}$$

The particular and homogeneous solution terms might be difficult to derive if the number of input and output variables is high and the number of boundary conditions is also high.

Also note that the Jacobian matrix for the multilayer perceptron will be affected if the problem includes boundary conditions. The approach is to make use of the product rule for differentiation in Equation (3.38)

Some examples of the use of boundary conditions for the multilayer perceptron can be found in Sections 5.1, 5.2, 5.3 and 5.4.

Lower and upper bounds

If some output variables are restricted to fall in some interval, then the problem is said to have lower and upper bounds. An easy way to treat lower and upper bounds is to post-process the outputs from Equation (3.38) in the next way,

$$\mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}) = \begin{cases} \inf(\mathbf{y}), & \mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}) < \inf(\mathbf{y}), \\ \mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}), & \inf(\mathbf{y}) \leq \mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}) \leq \sup(\mathbf{y}), \\ \sup(\mathbf{y}), & \mathbf{y}(\mathbf{x}; \underline{\alpha}, \underline{\beta}) > \sup(\mathbf{y}), \end{cases} \tag{3.43}$$

where $\inf(\mathbf{y})$ and $\sup(\mathbf{y})$ represent the infimum and supremum of the output vector $\mathbf{y}$, respectively.

Similarly, if some independent parameters are bounded they can be post-processed in the following manner,

$$\underline{\beta} = \begin{cases} \inf(\underline{\beta}), & \underline{\beta} < \inf(\underline{\beta}), \\ \underline{\beta}, & \inf(\underline{\beta}) \leq \underline{\beta} \leq \sup(\underline{\beta}), \\ \sup(\underline{\beta}), & \underline{\beta} > \sup(\underline{\beta}), \end{cases} \tag{3.44}$$

where $\inf(\underline{\beta})$ and $\sup(\underline{\beta})$ represent the infimum and supremum of the independent parameter vector $\underline{\beta}$, respectively.

Examples of problems with lower and upper bounds are found in Sections 6.2.1, or 8.2.1.

3.3.7 The input-output activity diagram

Following the contents of this section, an activity diagram for the input-output process in an extended class of multilayer perceptron with a single hidden layer of neurons can be drawn as in Figure 3.8.

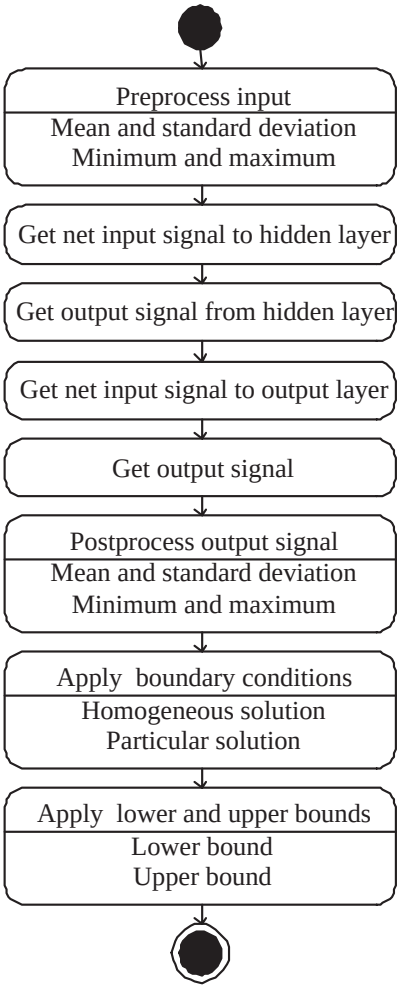


Figure 3.8: An activity diagram for the input-output process in the multilayer perceptron.

3.4 The objective functional

In order to perform a particular task a multilayer perceptron must be associated an objective functional. The objective functional defines the task the neural network is required to accomplish and provides a measure of the quality of the representation that the network is required to learn. The choice of a suitable objective functional for a multilayer perceptron depends on the variational problem at hand. The learning problem in the multilayer perceptron is thus formulated in terms of the minimization of the objective functional.

3.4.1 The variational problem

Traditionally, the learning problem for the multilayer perceptron has been formulated in terms of the minimization of an error function of the free parameters in the neural network, in order to fit the neural network to an input-target data set [15]. In that way, the only learning tasks allowed for the multilayer perceptron are data modeling type problems.

In a variational formulation for neural networks, the concept of error function, $e(\underline{\alpha})$, is changed by the concept or objective functional, $F[\mathbf{y}(\mathbf{x};\underline{\alpha})]$ [66]. An objective functional for the multilayer perceptron is of the form

$$\begin{array}{rccc} F: & V & \rightarrow & \mathbb{R} \\ & \mathbf{y}(\mathbf{x};\underline{\alpha}) & \mapsto & F[\mathbf{y}(\mathbf{x};\underline{\alpha})]. \end{array}$$

The objective functional defines the task the that the neural network is required to accomplish and provides a measure of the quality of the representation that the neural network is required to learn. In this way, the choice of a suitable objective functional depends on the particular application.

The learning problem for the multilayer perceptron can then be stated as the searching in the neural network function space for an element $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)$ at which the objective functional $F[\mathbf{y}(\mathbf{x};\underline{\alpha})]$ takes a maximum or a minimum value.

The tasks of maximization and minimization are trivially related to each other, since maximization of F is equivalent to minimization of $-F$, and vice versa. On the other hand, a minimum can be either a global minimum, the smallest value of the functional over its entire domain, or a local minimum, the smallest value of the functional within some local neighborhood.

As we will see, changing the concept of error function by the concept of objective functional allows us to extend the number of learning tasks for the multilayer perceptron to any variational problem. Some examples are optimal control problems [62], inverse problems [26] or optimal shape design problems [64].

The simplest variational problems are those in which no constraints are imposed on the solution. An unconstrained variational problem of historical interest is the brachistochrone problem [36]. However, there are many applications of the calculus of variations in which constraints are imposed. Such constraints are expressed as functionals. A classical constrained problem in the calculus of variations is the isoperimetric problem [36].

Unconstrained problems

The simplest variational problems for the multilayer perceptron are those in which no constraints are posed on the solution $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)$. In this way, the general unconstrained variational problem for the multilayer perceptron can be formulated as follows:

Problem 3 (Unconstrained variational problem)**.** Let V be the space of all functions $\mathbf{y}(\mathbf{x};\underline{\alpha})$ spanned by a multilayer perceptron, and let s be the dimension of V. Find a function $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in V$ for which the functional

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})],$$

defined on V, takes on a minimum value.

In other words, the unconstrained variational problem for the multilayer perceptron is stated in terms of the minimization of the objective functional associated to the neural network [66].

In Chapter 4 we use a multilayer perceptron to solve function regression and pattern recognition problems. That are unconstrained variational problems. The geodesic problem and the brachistochrone problem, which are solved in Chapter 5 are also examples of unconstrained variational problems.

Constrained problems

A variational problem for the multilayer perceptron can be specified by a set of constraints, which are equalities or inequalities that the solution $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)$ must satisfy. Such constraints are expressed as functionals. Thus, the general constrained variational problem for the multilayer perceptron can be formulated as follows:

Problem 4 (Constrained variational problem)**.** Let V be the space of all functions $\mathbf{y}(\mathbf{x};\underline{\alpha})$ spanned by a multilayer perceptron, and let s be the dimension of V. Find a function $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in V$ such that

$$C_i[\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)] = 0,$$

for $i = 1, \ldots, l$, and for which the functional

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})],$$

defined on V, takes on a minimum value.

In other words, the constrained variational problem for the multilayer perceptron consists of finding a vector of free parameters which makes all the constraints to be satisfied and the objective functional to be an extremum.

A common approach when solving a constrained variational problem is to reduce it into an unconstrained problem. This can be done by adding a penalty term to the objective functional for each of the constrains in the original problem. Adding a penalty term gives a large positive or negative value to the objective functional when infeasibility due to a constrain is encountered.

For the minimization case, the general constrained variational problem for the multilayer perceptron can be reformulated as follows:

Problem 5 (Reduced unconstrained variational problem)**.** Let V be the space consisting of all functions $\mathbf{y}(\mathbf{x};\underline{\alpha})$ that a given multilayer perceptron can define, and let s be the dimension of V. Find a function $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in V$ for which the functional

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})] + \sum_{i=1}^{l} \rho_i \|C_i[\mathbf{y}(\mathbf{x};\underline{\alpha})]\|^2,$$

defined on V and with $\rho_i > 0$, for $i = 1, \ldots, l$, takes on a minimum value.

The parameters ρ_i, for $i = 1, \ldots, l$, are called the penalty term weights, being l the number of constraints. Note that, while the squared norm of the constrained is the metric most used, any other suitable metric can be used.

For large values of ρ_i, it is clear that the solution $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)$ of Problem 5 will be in in a region where $C_i[\mathbf{y}(\mathbf{x};\underline{\alpha})]$, $i = 1, \ldots, l$, are small. Thus, for increasing values of ρ_i, it is expected that the the solution $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*)$ of Problem 5 will approach the constraints and, subject to being close, will minimize the objective functional $F[\mathbf{y}(\mathbf{x};\underline{\alpha})]$. Ideally then, as $\rho_i \to \infty$, $i = 1, \ldots, l$, the solution of Problem 5 will converge to the solution of Problem 4 [69].

In Chapter 5 we use a multilayer perceptron to solve the catenary problem and the isoperimetric problem. Other constrained problems are solved in Chapters 6 and 8.

3.4.2 The reduced function optimization problem

The objective functional, $F[\mathbf{y}(\mathbf{x};\underline{\alpha})]$, has an objective function associated, $f(\underline{\alpha})$, which is defined as a function of the free parameters in the neural network [66],

$$\begin{array}{rcl} f : \mathbb{R}^s & \to & \mathbb{R} \\ \underline{\alpha} & \mapsto & f(\underline{\alpha}). \end{array}$$

The objective function for the multilayer perceptron is represented as $f(\underline{\alpha})$, and can be visualized as a hypersurface, with $\alpha_1, \ldots, \alpha_s$ as coordinates, see Figure 3.9.

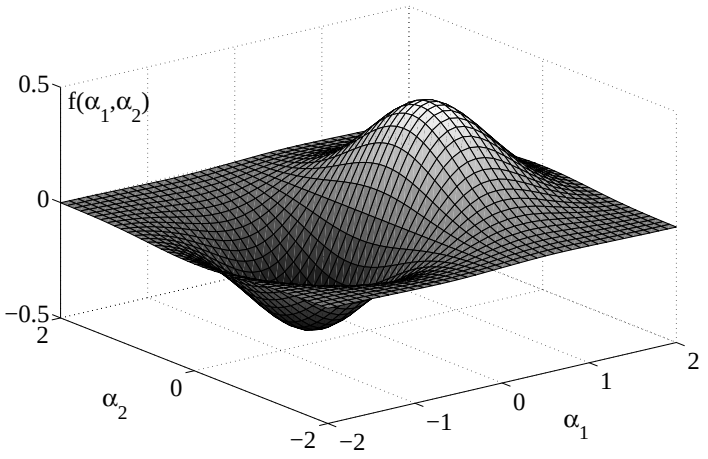


Figure 3.9: Geometrical representation of the objective function.

The minimum or maximum value of the objective functional is achieved for a vector of free parameters at which the objective function takes on a minimum or maximum value. Therefore, the learning problem in the multilayer perceptron, formulated as a variational problem, can be reduced to a function optimization problem [66].

Problem 6 (Reduced function optimization problem for the multilayer perceptron)**.** Let $\mathbb{R}^s$ be the space of all vectors $\underline{\alpha}$ spanned by the free parameters of a multilayer perceptron. Find a vector $\underline{\alpha}^* \in \mathbb{R}^s$ for which the function

$$f(\underline{\alpha}),$$

defined on $\mathbb{R}^s$, takes on a minimum value.

In this sense, a variational formulation for the multilayer perceptron provides a direct method for solving variational problems. The universal approximation properties for the multilayer perceptron cause neural computation to be a very appropriate paradigm for the solution of these problems.

3.4.3 The objective function gradient

We have seen that the objective functional $F[\mathbf{y}(\mathbf{x};\underline{\alpha})]$ of the multilayer perceptron has an objective function $f(\underline{\alpha})$ associated, which is defined as a function of the free parameters of the neural network; the learning problem in the multilayer perceptron is solved by finding the values of the free parameters which make the objective function to be an extremum.

The use of gradient information is of central importance in finding training algorithms which are sufficiently fast to be of practical use for large-scale applications. For a multilayer perceptron, the gradient vector ∇ of the objective function $f(\alpha_1, \ldots, \alpha_s)$ is written:

$$\nabla f(\alpha_1, \ldots, \alpha_s) = \left(\frac{\partial f}{\partial \alpha_1}, \ldots, \frac{\partial f}{\partial \alpha_s} \right). \tag{3.45}$$

Figure 3.10 represents the objective function gradient vector for the hypothetical case of a multilayer perceptron with two free parameters.

When the desired output of the multilayer perceptron for a given input is known, the objective function gradient can usually be found analytically using back-propagation. In some other circumstances exact evaluation of the gradient is not possible and numerical differentiation must be used.

The back-propagation algorithm for the calculus of the objective function gradient

Here we obtain the objective function gradient $\nabla f(\underline{\alpha})$ for the multilayer perceptron using the back-propagation algorithm. Results are derived for the one hidden layer perceptron shown in Figure 3.7, but they can be easily generalized to networks with several hidden layers of neurons.

In a one hidden layer perceptron, the net input signal for each neuron in the hidden layer is of the form

$$u_j^{(1)} = \sum_{i=0}^{n} \alpha_{ji}^{(1)} x_i, \tag{3.46}$$

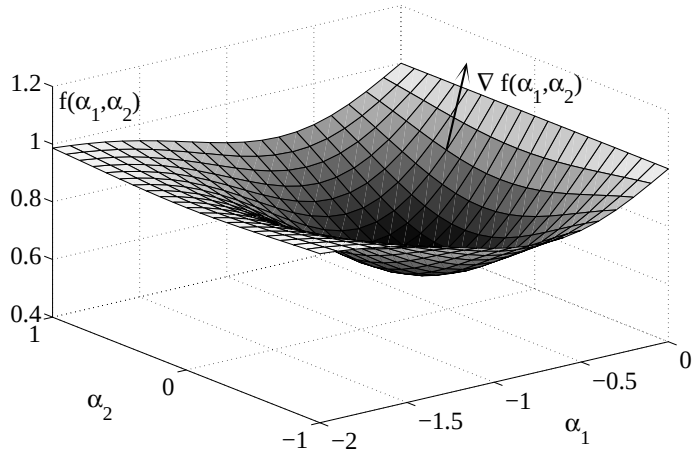


Figure 3.10: Illustration of the objective function gradient vector.

for $j = 1, \ldots, h_1$. In the same way, the net input signal for each neuron in the output layer is of the form

$$u_k^{(2)} = \sum_{j=0}^{h_1} \alpha_{kj}^{(2)} y_j^{(1)}, \tag{3.47}$$

for $k = 1, \ldots, m$.

The net input signal of neuron j in the hidden layer is transformed by an activation function $g^{(1)}$ to give the activation $y_j^{(1)}$

$$y_j^{(1)} = g^{(1)}(u_j^{(1)}), \tag{3.48}$$

for $j = 1, \ldots, h_l$. Similarly, for the output layer,

$$y_k = g^{(2)}(u_k^{(2)}), \tag{3.49}$$

for $k = 1, \ldots, m$. As we said before, the learning algorithm will aim to find suitable values for the free parameters in the neural network by optimizing the objective function $f(\underline{\alpha})$. On the other hand, we consider objective functions which ca be expressed as a differentiable function of the output signals $\mathbf{y}$ from the neural network. Therefore, the objective function of a multilayer perceptron can be expressed as a function of the output signals, which depend in turn on the free parameters

$$f = f(\alpha_1, \ldots, \alpha_s; y_1, \ldots, y_m). \tag{3.50}$$

We want to find a procedure to evaluate the derivatives of the objective function f with respect to the free parameters in the network $\alpha_1, \ldots, \alpha_s$.

First we obtain the activation of all hidden and output neurons by a consecutive applying of (3.46), (3.48), (3.47) and (3.49). This process is often called feed-propagation, since it can be considered as a feed flow of information through the neural network.

Consider now the evaluation of the derivative of the objective function f with respect to some free parameter α_{ji}. f depends on the free parameter α_{ji} only though the net input signal u_j to neuron j. We can then apply the chain rule for partial derivatives to give, for the hidden layer

$$\frac{\partial f}{\partial \alpha_{ji}^{(1)}} = \frac{\partial f}{\partial u_j^{(1)}} \frac{\partial u_j^{(1)}}{\partial \alpha_{ji}^{(1)}}, \tag{3.51}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, h_1$. Likewise, for the output layer,

$$\frac{\partial f}{\partial \alpha_{kj}^{(2)}} = \frac{\partial f}{\partial u_k^{(2)}} \frac{\partial u_k^{(2)}}{\partial \alpha_{kj}^{(2)}}, \tag{3.52}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. We now introduce the notation

$$\delta \equiv \frac{\partial f}{\partial u}. \tag{3.53}$$

The quantity δ is called back-propagation error, and it is considered for each neuron in the neural network. Therefore, the definition of the back-propagation error for a neuron in the hidden layer is

$$\delta_j^{(1)} \equiv \frac{\partial f}{\partial u_j^{(1)}}, \tag{3.54}$$

for $j = 1, \ldots, h_1$. The definition of the back-propagation error for a neuron in the output layer is

$$\delta_k^{(2)} \equiv \frac{\partial f}{\partial u_k^{(2)}}, \tag{3.55}$$

for $k = 1, \ldots, m$. On the other hand, using Equation (3.46) we can write, for the hidden layer

$$\frac{\partial u_j^{(1)}}{\partial \alpha_{ji}^{(1)}} = x_i, \tag{3.56}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, h_1$. Similarly, using Equation (3.47) we can write, for the output layer,

$$\frac{\partial u_k^{(2)}}{\partial \alpha_{kj}^{(2)}} = y_j^{(1)}, \tag{3.57}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. Substituting (3.54) and (3.56) in (3.51) we obtain, for the hidden layer

$$\frac{\partial f}{\partial \alpha_{ji}^{(1)}} = \delta_j^{(1)} x_i, \tag{3.58}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, h_1$. Likewise, substituting (3.55) and (3.57) in (3.52) we obtain, for the output layer

$$\frac{\partial f}{\partial \alpha_{kj}^{(2)}} = \delta_k^{(2)} y_j^{(1)}, \tag{3.59}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. Thus, to evaluate the required derivatives, we only need to evaluate the value of δ for each neuron in the neural network and then to apply (3.54) and (3.55).

The evaluation of $\delta_k^{(2)}$ for the output neurons is quite simple. From the definition in Equation (3.54) we have

$$\begin{aligned} \delta_k^{(2)} &\equiv \frac{\partial f}{\partial u_k^{(2)}} \\ &= \frac{\partial f}{\partial y_k^{(2)}} \frac{\partial {y_k}^{(2)}}{\partial u_k^{(2)}} \\ &= g'^{(2)}(u_k^{(2)}) \frac{\partial f}{\partial y_k^{(2)}}, \end{aligned} \tag{3.60}$$

for $k = 1, \ldots, m$. To evaluate $\delta_j^{(1)}$ for the hidden neurons we make use of the chain rule for partial derivatives,

$$\begin{aligned} \delta_j^{(1)} &\equiv \frac{\partial f}{\partial u_j^{(1)}} \\ &= \sum_{k=1}^{m} \frac{\partial f}{\partial u_k^{(2)}} \frac{\partial u_k^{(2)}}{\partial u_j^{(1)}}, \end{aligned} \tag{3.61}$$

for $j = 1, \ldots, h_1$. Substituting now the definition of δ for the neurons in the output layer given in (3.55) into (3.61), and making use of (3.46) and (3.48), we obtain

$$\delta_j^{(1)} = g'^{(1)}(u_j^{(1)}) \sum_{k=1}^{m} \alpha_{kj}^{(2)} \delta_k^{(2)}, \tag{3.62}$$

for $j = 1, \ldots, h_1$. We can summarize the back-propagation procedure to evaluate the derivatives of the objective function with respect to the free parameters in just four steps:

1. Apply an input vector $\mathbf{x}$ to the neural network and feed-propagate it using (3.46), (3.48), (3.47) and (3.49) to find the activation of all hidden and output neurons.

2. Evaluate the errors $\delta_k^{(2)}$ for all output neurons using (3.60).

3. Back-propagate the errors $\delta_k^{(2)}$ by using (3.62) to obtain $\delta_j^{(1)}$ for each hidden neuron in the neural network.

4. Use (3.54) and (3.55) to evaluate the required derivatives of the objective function with respect to the free parameters in the hidden and output layers, respectively.

A C++ implementation of this algorithm for some particular objective functions can be found in the classes SumSquaredError, MeanSquaredError, RootMeanSquaredError, NormalizedSquaredError, MinkowskiError and RegularizedMeanSquaredError of Flood [61].

In Sections 4.2 and 4.3 we make use of the back-propagation algorithm to obtain the objective function gradient of the sum squared error, which is a common objective functional for function regression and pattern recognition problems.

Numerical differentiation for the calculus of the objective function gradient

There are many applications when it is not possible to obtain the objective function gradient $\nabla f(\underline{\alpha})$ using the back-propagation algorithm, and it needs to be computed numerically. This can be done by perturbing each free parameter in turn, and approximating the derivatives by using the finite differences method

$$\frac{\partial f}{\partial \alpha_i} = \frac{f(\alpha_i + \epsilon) - f(\alpha_i)}{\epsilon} + \mathcal{O}(\epsilon), \tag{3.63}$$

for $i = 1, \ldots, s$ and for some small numerical value of ϵ.

The accuracy of the finite differences method can be improved significantly by using symmetrical central differences of the form

$$\frac{\partial f}{\partial \alpha_i} = \frac{f(\alpha_i + \epsilon) - f(\alpha_i - \epsilon)}{2\epsilon} + \mathcal{O}(\epsilon^2), \tag{3.64}$$

also for $i = 1, \ldots, s$ and for some small numerical value of ϵ.

We have seen that the derivatives of the objective function with respect to the free parameters in the hidden layer can be expressed efficiently through the relation

$$\frac{\partial f}{\partial \alpha_{ji}^{(1)}} = \frac{\partial f}{\partial u_j^{(1)}} x_i, \tag{3.65}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, h_1$. Similarly, the derivatives of the objective function with respect to the free parameters in the output layer are given by

$$\frac{\partial f}{\partial \alpha_{kj}^{(2)}} = \frac{\partial f}{\partial u_k^{(2)}} y_j^{(1)}, \tag{3.66}$$

for $j = 1, \ldots, h_1$ and $k = 1, \ldots, m$. Instead of using the technique of central differences to evaluate the derivatives $\partial f / \partial \alpha$ directly, we can use it to estimate $\partial f / \partial u$ since

$$\frac{\partial f}{\partial u_i} = \frac{f(u_i + \epsilon) - f(u_i - \epsilon)}{2\epsilon} + \mathcal{O}(\epsilon^2), \tag{3.67}$$

for $i = 1, \ldots, r$, where $r = h_1 + m$ is the total number of neurons in the network. We can then make use of Equations (3.65) and (3.66) to evaluate the required derivatives.

In Sections 5.2 and 5.4 we make use of numerical differentiation to obtain the objective function gradient of the objective functionals for the brachistochrone and the isoperimetric problems, respectively.

In a software implementation, when possible, derivatives of the objective function f with respect to the free parameters in the network $\underline{\alpha}$ should be evaluated using back-propagation, since this provides the greatest accuracy and numerical efficiency.

The source code for the calculus of the gradient vector by means of central differences is within the class ObjectiveFunctional of Flood [61].

3.4.4 The objective function Hessian

There are some training algorithms which also make use of the Hessian matrix of the objective function to search for an optimal set of free parameters. The Hessian matrix of the objective function is written

$$\mathbf{H}f(\underline{\alpha}) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial \alpha_1^2} & \cdots & \dfrac{\partial^2 f}{\partial \alpha_1 \partial \alpha_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial \alpha_n \partial \alpha_1} & \cdots & \dfrac{\partial^2 f}{\partial \alpha_n^2} \end{pmatrix} \tag{3.68}$$

The most general scheme to calculate the Hessian matrix is to apply numerical differentiation. However, there are some objective functions which have an analytical solution for the Hessian matrix, which can be calculated using a back-propagation algorithm [15].

Numerical differentiation for the calculus of the objective function Hessian

As it happens for the gradient vector, there are many applications when analytical evaluation of the Hessian is not possible, and it must be computed numerically. This can be done by perturbing each argument element in turn, and approximating the derivatives by using the central differences method

$$\begin{aligned} H_{ij} &= \frac{\partial^2 f}{\partial \alpha_i \partial \alpha_j} \\ &= \frac{f(\alpha_i + \epsilon, \alpha_j + \epsilon)}{4\epsilon^2} - \frac{f(\alpha_i + \epsilon, \alpha_j - \epsilon)}{4\epsilon^2} \\ &- \frac{f(\alpha_i - \epsilon, \alpha_j + \epsilon)}{4\epsilon^2} + \frac{f(\alpha_i - \epsilon, \alpha_j - \epsilon)}{4\epsilon^2} + \mathcal{O}(\epsilon^2), \end{aligned} \tag{3.69}$$

for $i, j = 1, \ldots, s$, and where s is the number of free parameters in the neural network.

Equation (3.69) is implemented in the class ObjectiveFunctional of Flood [61].

3.5 The training algorithm

The procedure used to carry out the learning process in a neural network is called the training algorithm. There are many different training algorithms for the multilayer perceptron. Some of

the most used are the quasi-Newton method or the evolutionary algorithm.

As we said in Chapter 3.4, the learning problem in the multilayer perceptron, formulated as a variational problem, can be reduced to a function optimization problem. The training algorithm is entrusted to solve the reduced function optimization problem, by adjusting the parameters in the neural network so as to optimize the objective function.

3.5.1 The function optimization problem

The learning problem in the multilayer perceptron has been reduced to the searching in an s-dimensional space for a parameter vector $\underline{\alpha}^*$ for which the objective function f takes a maximum or a minimum value.

The tasks of maximization and minimization are trivially related to each other, since maximization of $f(\underline{\alpha})$ is equivalent to minimization of $-f(\underline{\alpha})$, and vice versa.

On the other side, a minimum can be either a global minimum, the smallest value of the function over its entire range, or a local minimum, the smallest value of the function within some local neighborhood. Finding a global minimum is, in general, a very difficult problem [101].

Consider an objective function

$$\begin{aligned} f \quad &: \quad \mathbb{R}^s \to \mathbb{R} \\ &\quad\; \underline{\alpha} \quad \mapsto f(\underline{\alpha}) \end{aligned}$$

continuous, derivable, and with continuous derivatives.

The global minimum condition for the parameter vector $\underline{\alpha}^*$ is stated as

$$f(\underline{\alpha}^*) \leq f(\underline{\alpha}), \tag{3.70}$$

for all $\alpha_1, \ldots, \alpha_s$ in the parameter space.

On the other hand, the necessary local minimum condition for the parameter vector $\underline{\alpha}^*$ is stated as

$$\nabla f(\underline{\alpha}^*) = 0. \tag{3.71}$$

The objective function is, in general, a non linear function of the parameters. As a consequence, it is not possible to find closed training algorithms for the minima. Instead, we consider a search through the parameter space consisting of a succession of steps of the form

$$\alpha^{(i+1)} = \alpha^{(i)} + \Delta\alpha^{(i)}, \tag{3.72}$$

where i labels the iteration step, or epoch. The quantity $\Delta\alpha^{(i)}$ is called parameter vector increment. Different training algorithms involve different choices for this vector.

In this way, to train a multilayer perceptron we start with an initial estimation of the parameter vector $\underline{\alpha}^{(0)}$ (often chosen at random) and we generate a sequence of parameter vectors $\underline{\alpha}^{(1)}$, $\underline{\alpha}^{(2)}, \ldots$, so that the objective function f is reduced at each iteration of the algorithm, that is

$$f(\alpha^{(i+1)}) < f(\alpha^{(i)}). \tag{3.73}$$

The training algorithm stops when a specified condition is satisfied. Some stopping criteria commonly used are [27]:

1. A maximum number of epochs is reached.
2. A maximum amount of computing time has been exceeded.
3. Evaluation has been minimized to a goal value.
4. Evaluation improvement in one epoch is less than a set value.
5. The norm of the objective function gradient falls below a goal.

Figure 3.11 is a state diagram of this iterative procedure, showing states and transitions in the training process of a multilayer perceptron.

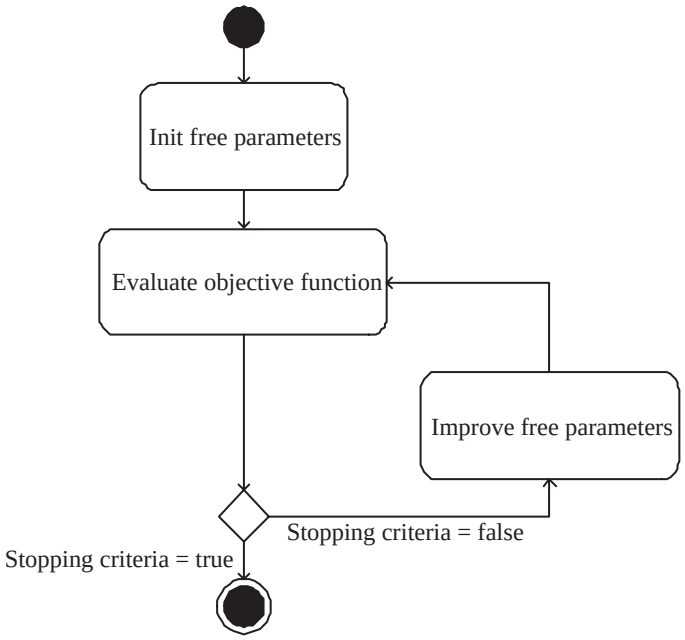


Figure 3.11: Training process in the multilayer perceptron.

The training process is determined by the way in which the adjustment of the parameters in the neural network takes place. There are many different training algorithms, which have a variety of different computation and storage requirements. Moreover, there is not a training algorithm best suited to all locations [101].

Training algorithms might require information from the objective function only, the gradient vector of the objective function or the Hessian matrix of the objective function [83]. These methods, in turn, can perform either global or local optimization.

Zero-order training algorithms make use of the objective function only. The most significant zero-order training algorithms are stochastic, which involve randomness in the optimization process. Examples of these are random search and evolutionary algorithms[39] [34] or particle swarm optimization [52], which are global optimization methods .

First-order training algorithms use the objective function and its gradient vector [11]. Examples of these are gradient descent methods, conjugate gradient methods, scaled conjugate gradient

methods [75] or quasi-Newton methods. Gradient descent, conjugate gradient, scaled conjugate gradient and quasi-Newton methods are local optimization methods [69].

Second-order training algorithms make use of the objective function, its gradient vector and its Hessian matrix [11]. Examples for second-order methods are Newton's method and the Levenberg-Marquardt algorithm [40]. Both of them are local optimization methods [69].

3.5.2 Random search

Random search is the simplest possible training algorithm for the multilayer perceptron. It is a stochastic method which requires information from the objective function only, and therefore a zero order optimization method.

The random search method simply consists of sampling a stream parameter vectors

$$\underline{\alpha}^{(1)}, \underline{\alpha}^{(2)}, \ldots$$

distributed at random, and while evaluating their objective function

$$f^{(1)}(\underline{\alpha}^{(1)}), f^{(2)}(\underline{\alpha}^{(2)}), \ldots$$

until a stopping criterium is satisfied.

Random search can be performed either using a uniform distribution, i.e., with constant probability, or using a normal distribution, i.e., with probability defined by a mean and an standard deviation.

Unfortunately, convergence is extremely slow in most cases, so this training algorithm is only in practice used to obtain a good initial guess for other more efficient methods. A C++ implementation of random search can be found in the class RandomSearch of Flood [61].

3.5.3 Gradient descent

Gradient descent, sometimes also known as steepest descent, is a local method which requires information from the gradient vector, and hence it is a first order training algorithm. It acts in a deterministic manner.

The method begins at a point $\underline{\alpha}^{(0)}$ and, until a stopping criterium is satisfied, moves from $\underline{\alpha}^{(i)}$ to $\underline{\alpha}^{(i+1)}$ along the line extending from $\underline{\alpha}^{(i)}$ in the direction of $-\nabla f(\underline{\alpha}^{(i)})$, the local downhill gradient. The gradient vector of the objective function for the multilayer perceptron is described in Section 3.4.3.

Let denote $\mathbf{g} \equiv \nabla f(\underline{\alpha})$ the gradient vector. Therefore, starting from a parameter vector $\underline{\alpha}^{(0)}$, the gradient descent method takes the form of iterating

$$\underline{\alpha}^{(i+1)} = \underline{\alpha}^{(i)} - \eta^{(i)} \mathbf{g}^{(i)}, \tag{3.74}$$

for $i = 0, 1, \ldots$, and where the parameter $\eta^{(i)}$ is called the training rate. This value can either set to a fixed value or found by line minimization along the train direction at each epoch. Provided that the train rate is well chosen, the value of f will decrease at each successive step, eventually reaching to vector of parameters $\underline{\alpha}^*$ at which some stopping criterium is satisfied.

The choose of a suitable value for a fixed train rate presents a serious difficulty. If η is too large, the algorithm may overshoot leading to an increase in f and possibly to divergent oscillations,

resulting in a complete breakdown in the algorithm. Conversely, if η is chosen to be very small the search can proceed extremely slowly, leading to long computation times. Furthermore, a good value for η will typically change during the course of training [15].

For that reason, an optimal value for the train rate obtained by line minimization at each successive epoch is generally preferable. Here a search is made along the train direction to determine the optimal train rate, which minimizes the objective function along that line. Section 3.5.7 describes different one-dimensional minimization algorithms.

Figure 3.12 is a state diagram for the training process of a neural network with gradient descent. Improvement of the parameters is performed by obtaining first the gradient descent train direction and then a suitable training rate.

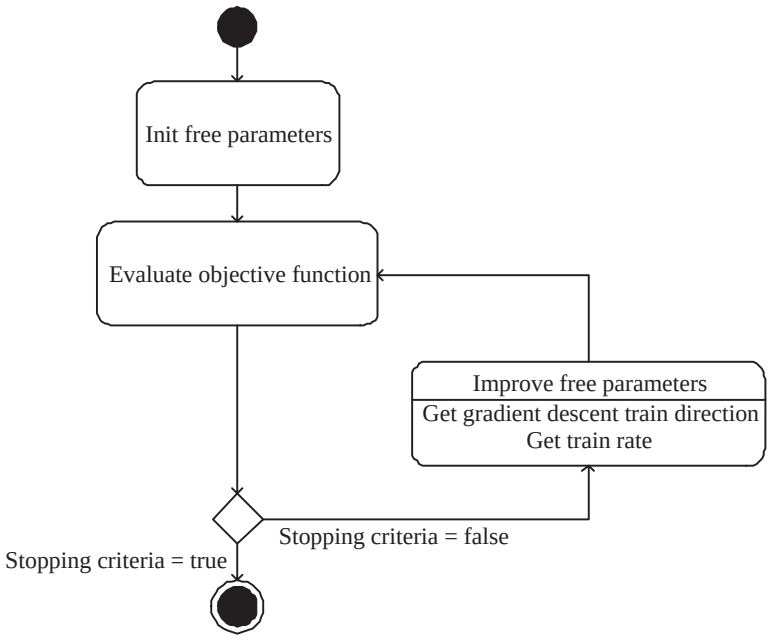


Figure 3.12: Training process with the gradient descent training algorithm.

The gradient descent training algorithm has the severe drawback of requiring many iterations for functions which have long, narrow valley structures. Indeed, the local downhill gradient is the direction in which the objective function decreases most rapidly, but this does not necessarily produce the fastest convergence. See [69] for a detailed discussion of this optimization method. The class GradientDescent within Flood [61] contains a C++ source of the gradient descent training algorithm.

3.5.4 Newton's method

The Newton's method is a class of local algorithm which makes use of the Hessian matrix of the objective function. In this way it is a second order method. On the other hand, the Newton's method behaves in a deterministic fashion.

Consider the quadratic approximation of f at $\underline{\alpha} = \underline{\alpha}^{(i)}$ using the Taylor's series expansion

$$f(\underline{\alpha}) = f(\underline{\alpha}^{(i)}) + \nabla f(\underline{\alpha} - \underline{\alpha}^{(i)}) + \frac{1}{2}(\underline{\alpha} - \underline{\alpha}^{(i)})^{\mathrm{T}} \cdot \mathbf{H}f(\underline{\alpha}^{(i)}) \cdot (\underline{\alpha} - \underline{\alpha}^{(i)}), \tag{3.75}$$

where $\mathbf{H}f(\underline{\alpha}^{(i)})$ is the Hessian matrix of f evaluated at the point $\underline{\alpha}^{(i)}$. The Hessian matrix of the objective function for the multilayer perceptron is described in Section 3.4.4.

By setting $\nabla f(\underline{\alpha})$ in Equation (3.75) equal to $\mathbf{0}$ for the minimum of $f(\underline{\alpha})$, we obtain

$$\begin{aligned} \nabla f(\underline{\alpha}) &= \nabla f(\underline{\alpha}^{(i)}) + \mathbf{H}f(\underline{\alpha}^{(i)}) \cdot (\underline{\alpha} - \underline{\alpha}^{(i)}) \\ &= \mathbf{0}. \end{aligned} \tag{3.76}$$

If $\mathbf{H}f(\underline{\alpha}^{(i)})$ is not singular, Equation (3.76) leads to an expression for the location of the minimum of the objective function,

$$\underline{\alpha}^{*} = \underline{\alpha}^{(i)} - \mathbf{H}^{-1}f(\underline{\alpha}^{(i)}) \cdot \nabla f(\underline{\alpha}^{(i)}), \tag{3.77}$$

where $\mathbf{H}^{-1}f(\underline{\alpha}^{(i)})$ is the inverse of the Hessian matrix of f evaluated at the point $\underline{\alpha}^{(i)}$.

Equation (3.77) would be exact for a quadratic objective function. However, since higher order terms have been neglected, this is to be used iteratively to find the optimal solution $\underline{\alpha}^{*}$. Let denote $\mathbf{g} \equiv \nabla f(\underline{\alpha})$ and $\mathbf{H} \equiv \mathbf{H}f(\underline{\alpha})$. Therefore, starting from a parameter vector $\underline{\alpha}^{(0)}$, the iterative formula for the Newton's method can be written

$$\underline{\alpha}^{(i+1)} = \underline{\alpha}^{(i)} - \mathbf{H}^{-1(i)}\mathbf{g}^{(i)}, \tag{3.78}$$

for $i = 0, 1, \ldots$ and until some stopping criterium is satisfied.

The vector $\mathbf{H}^{-1}\mathbf{g}$ is known as the Newton's increment. But note that this increment for the parameters may move towards a maximum or a saddle point rather than a minimum. This occurs if the Hessian is not positive definite, so that there exist directions of negative curvature. Thus, the objective function evaluation is not guaranteed to be reduced at each iteration. Moreover, the Newton's increment may be sufficiently large that it takes us outside the range of validity of the quadratic approximation. In this case the algorithm could become unstable.

In order to prevent such troubles, the Newton's method in Equation (3.5.4) is usually modified as

$$\underline{\alpha}^{(i+1)} = \underline{\alpha}^{(i)} - \eta^{(i)}\mathbf{H}^{-1(i)}\mathbf{g}^{(i)}, \tag{3.79}$$

where the training rate $\eta^{(i)}$ can either set to a fixed value or found by line minimization. See Section 3.5.7 for a description of several one-dimensional minimization algorithms.

In Equation (3.79), the vector $\mathbf{H}^{-1}\mathbf{g}$ is now called the Newton's train direction. The sequence of points $\underline{\alpha}^{(0)}, \underline{\alpha}^{(1)}, \ldots$ can be shown here to converge to the actual solution $\underline{\alpha}^{*}$ from any initial point $\underline{\alpha}^{(0)}$ sufficiently close the to solution, and provided that $\mathbf{H}$ is nonsingular.

The state diagram for the training process with the Newton's method is depicted in Figure 3.12. Here improvement of the parameters is performed by obtaining first the Newton's method train direction and then a suitable training rate.

There are still several difficulties with such an approach, however. First, an exact evaluation of the Hessian matrix is computationally demanding. This evaluation would be prohibitively

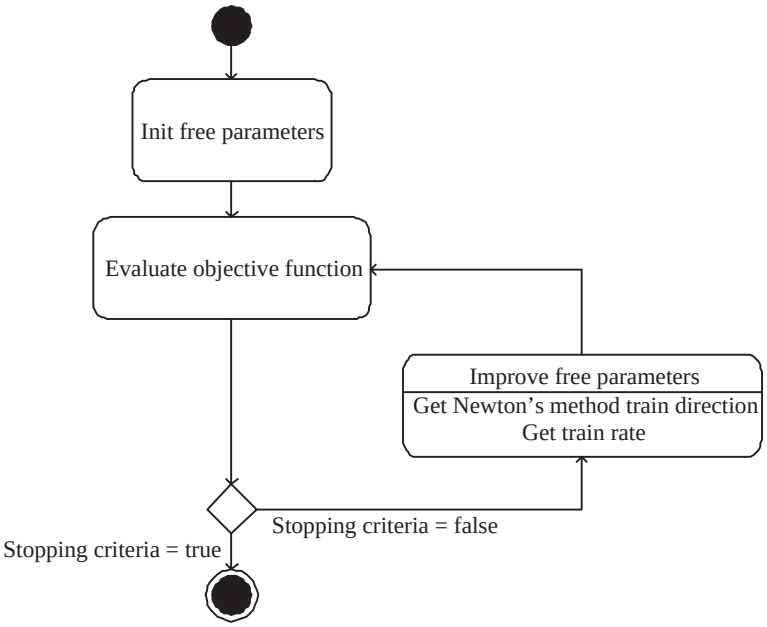


Figure 3.13: Training process with the Newton's method.

expensive if done at each stage of an iterative algorithm. Second, the Hessian must be inverted, and so is also computationally demanding. In [69] a complete description of the Newton's method can be found. Also, a C++ implementation of this training algorithm is included within the class NewtonMethod of Flood [61]

3.5.5 Conjugate gradient

Conjugate gradient is a local algorithm for an objective function whose gradient can be computed, belonging for that reason to the class of first order methods. According to its behavior, it can be described as a deterministic method.

The conjugate gradient method can be regarded as being somewhat intermediate between the method of gradient descent and Newton's method [69]. It is motivated by the desire to accelerate the typically slow convergence associated with gradient descent while avoiding the information requirements associated with the evaluation, storage, and inversion of the Hessian matrix as required by the Newton's method. In the conjugate gradient algorithm search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions [27]. These train directions are conjugated with respect to the Hessian matrix. A set of vectors $\mathbf{d}_k$ are said to be conjugated with respect to the matrix $\mathbf{H}$ if only if

$$\mathbf{d}_i^{\mathrm{T}} \mathbf{H} \mathbf{d}_j = 0, \tag{3.80}$$

for all $i \neq j$ and provided that $\mathbf{H}$ is not singular. An elemental property of a set of conjugate directions is that these vectors are linearly independent. Therefore, if the number of parameters is s, the maximum size of a set of conjugate directions is also s.

Let denote $\mathbf{g} \equiv \nabla f(\underline{\alpha})$ and $\mathbf{h}$ the train direction vector. Then, starting with an initial parameter vector $\underline{\alpha}^{(0)}$ and an initial train direction vector $\mathbf{h}^{(0)} = -\mathbf{g}^{(0)}$, the conjugate gradient method constructs a sequence of train directions from the recurrence

$$\mathbf{h}^{(i+1)} = \mathbf{g}^{(i+1)} + \gamma^{(i)}\mathbf{h}^{(i)}, \tag{3.81}$$

for $i = 0, 1, \ldots$.

The parameters can then be improved according to the formula

$$\underline{\alpha}^{(i+1)} = \underline{\alpha}^{(i)} + \eta^{(i)}\mathbf{h}^{(i)}, \tag{3.82}$$

also for $i = 0, 1, \ldots$, and where $\eta^{(i)}$ is the train rate, which is usually found by line minimization.

The various versions of conjugate gradient are distinguished by the manner in which the parameter $\gamma^{(i)}$ is constructed. For the Fletcher-Reeves update the procedure is [33]

$$\gamma_{FR}^{(i)} = \frac{\mathbf{g}^{(i+1)}\mathbf{g}^{(i+1)}}{\mathbf{g}^{(i)}\mathbf{g}^{(i)}}, \tag{3.83}$$

where $\gamma_{FR}^{(i)}$ is called the Fletcher-Reeves parameter.

For the Polak-Ribiere update the procedure is

$$\gamma_{PR}^{(i)} = \frac{(\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\mathbf{g}^{(i+1)}}{\mathbf{g}^{(i)}\mathbf{g}^{(i)}}, \tag{3.84}$$

where $\gamma_{PR}^{(i)}$ is called the Polak-Ribiere parameter.

It can be shown that both the Fletcher-Reeves and the Polak-Ribiere train directions indeed satisfy Equation (3.82).

There is some evidence that the Polak-Ribiere formula accomplishes the transition to further iterations more efficiently: When it runs out of steam, it tends to reset the train direction $\mathbf{h}$ to be down the local gradient, which is equivalent to beginning the conjugate-gradient procedure again [83].

For all conjugate gradient algorithms, the train direction is periodically reset to the negative of the gradient. The standard reset point occurs every s epochs, being s is the number of parameters in the multilayer perceptron [82].

Figure 3.14 is a state diagram for the training process with the conjugate gradient. Here improvement of the parameters is done by first computing the conjugate gradient train direction and then a suitable train rate in that direction.

Conjugate gradient methods have proved to more effective than gradient descent or the Newton's method in dealing with general objective functions. A detailed discussion of the conjugate gradient method can be found in [69]. The library Flood also implements this method in the C++ class ConjugateGradient [61].

3.5.6 Quasi-Newton method

The quasi-Newton method can be classified as a local, first order and deterministic training algorithm for the multilayer perceptron.

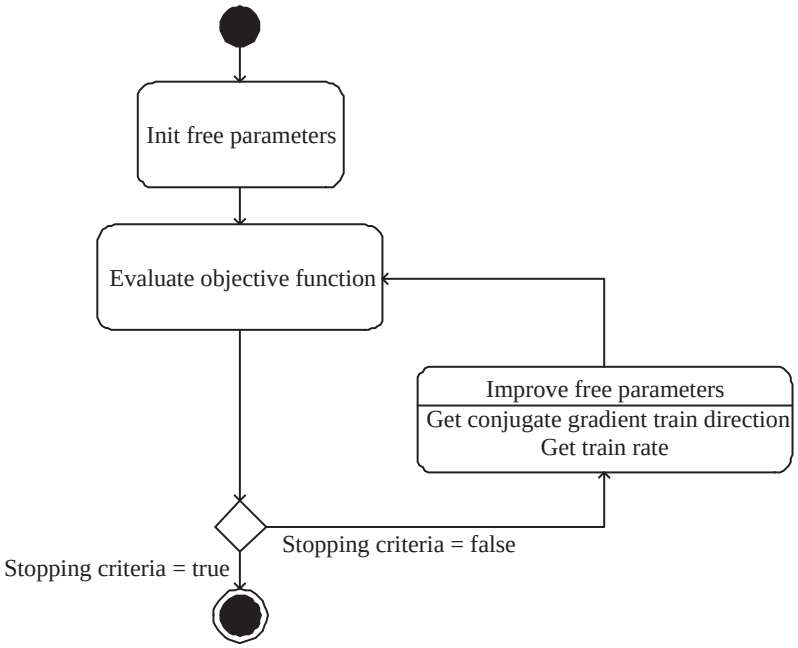


Figure 3.14: Training process with the conjugate gradient training algorithm.

In Section 3.5.4 it was argued that a direct application of the Newton's method, as given by Equation (3.79), would be computationally prohibitive since it would require too many operations to evaluate the Hessian matrix and compute its inverse. Alternative approaches, known as quasi-Newton or variable metric methods, are based on Equation (3.79), but instead of calculating the Hessian directly, and then evaluating its inverse, they build up an approximation to the inverse Hessian over a number of steps.

The Hessian matrix is composed of the second partial derivatives of the objective function. The basic idea behind the quasi-Newton or variable metric methods is to approximate $\mathbf{H}^{-1}$ by another matrix $\mathbf{G}$, using only the first partial derivatives of f. If $\mathbf{H}^{-1}$ is approximated by $\mathbf{G}$, the Newton formula (3.79) can be expressed as

$$\underline{\alpha}^{(i+1)} = \underline{\alpha}^{(i)} - \eta^{(i)}\mathbf{G}^{(i)}\mathbf{g}^{(i)}, \tag{3.85}$$

where $\eta^{(i)}$ can either set to a fixed value or found by line minimization.

Implementation of Equation (3.85) involves generating a sequence of matrices $\mathbf{G}$ which represent increasingly accurate approximation to the inverse Hessian $\mathbf{H}^{-1}$, sing only information on the first derivatives of the objective function. The problems arising from Hessian matrices which are not positive definite are solved by starting from a positive definite matrix (such as the unit matrix) and ensuring that the update procedure is such that the approximation to the inverse Hessian is guaranteed to remain positive definite. The approximation $\mathbf{G}$ of the inverse Hessian must be constructed so as to satisfy this condition also.

The two most commonly used update formulae are the Davidon-Fletcher-Powel (DFP) algorithm (sometimes referred to as simply Fletcher-Powel (FP) algorithm) and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm.

The DFP algorithm is given by

$$\begin{aligned}\mathbf{G}^{(i+1)} &= \mathbf{G}^{(i)} \\ &+ \frac{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)}) \otimes (\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)})}{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)}) \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})} \\ &+ \frac{\left[\mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\right] \otimes \left[\mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\right]}{(\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)}) \cdot \mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})},\end{aligned} \tag{3.86}$$

where $\otimes$ denotes the outer or direct product of two vectors, which is a matrix: The ij component of $\mathbf{u} \otimes \mathbf{v}$ is $u_i v_j$.

The BFGS algorithm is exactly the same, but with one additional term

$$\begin{aligned}\mathbf{G}^{(i+1)} &= \mathbf{G}^{(i)} \\ &+ \frac{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)}) \otimes (\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)})}{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)}) \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})} \\ &+ \frac{\left[\mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\right] \otimes \left[\mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\right]}{(\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)}) \cdot \mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})} \\ &+ \left[(\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)}) \cdot \mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})\right] \mathbf{u} \otimes \mathbf{u},\end{aligned} \tag{3.87}$$

where the vector $\mathbf{u}$ is given by

$$\begin{aligned}\mathbf{u} &= \frac{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)})}{(\underline{\alpha}^{(i+1)} - \underline{\alpha}^{(i)}) \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})} \\ &- \frac{\mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})}{(\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)}) \cdot \mathbf{G}^{(i)} \cdot (\mathbf{g}^{(i+1)} - \mathbf{g}^{(i)})}.\end{aligned} \tag{3.88}$$

It has become generally recognized that, empirically, the BFGS scheme is superior than the DFP scheme [83].

A state diagram of the training process of a multilayer perceptron is depicted in Figure 3.15. Improvement of the parameters is performed by first obtaining the quasi-Newton train direction and then finding a satisfactory train rate.

The quasi-Newton method is the algorithm of choice in most of the applications included in this PhD Thesis. This is discussed in detail in [69]. A software implementation of the quasi-Newton method in the C++ programming language can be found in the QuasiNewtonMethod class of Flood [61].

3.5.7 One-dimensional minimization algorithms

The basic schema of most the training algorithms for the multilayer perceptron is to produce a sequence of improved approximations to the optimal parameter vector according to the following basis:

1. Start with an initial trial vector $\underline{\alpha}^{(0)}$.
2. Find a suitable train direction $\mathbf{d}^{(0)}$ in the general direction of the optimum.
3. Find an appropriate train rate $\eta^{(0)}$ for movement along that train direction.

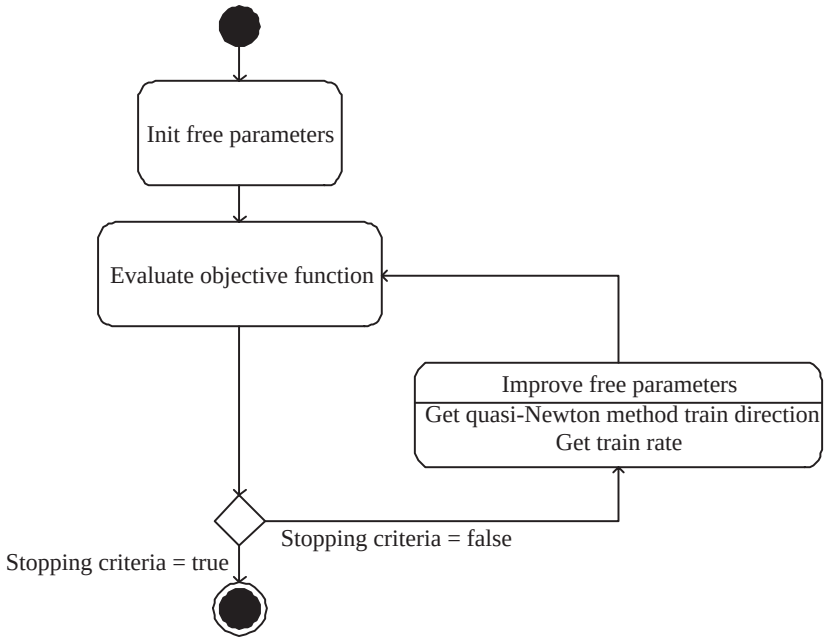


Figure 3.15: Training process with the quasi-Newton method.

4. Obtain the new approximation $\underline{\alpha}^{(1)}$.

5. Test whether any stopping criteria is satisfied. Otherwise, repeat step 2 onward.

The train rate in step 3 can either set to a fixed value or found by line minimization along the train direction at each epoch. In general, an optimal value for the train rate obtained by line minimization at each successive epoch is preferable.

Line minimization algorithms begin by locating an interval in which the minimum of the objective function along occurs. A minimum is known to be bracketed when there is a triplet of points $a < b < c$ such that $f(a) > f(b) < f(c)$. In this case we know that f has a minimum in the interval (a, c).

The golden section method brackets that minimum until the distance between the two outer points in the bracket is less than a defined tolerance [83].

The Brent's method performs a parabolic interpolation until the distance between the two outer points defining the parabola is less than a tolerance [83].

Flood implements the golden section and the Brent's method inside the C++ classes GradientDescent, NewtonMethod, ConjugateGradient and QuasiNewtonMethod of Flood [61].

3.5.8 Evolutionary algorithm

A global training algorithm for the multilayer perceptron is the evolutionary algorithm, a class of which is the genetic algorithm. This is a stochastic method based on the mechanics of natural genetics and biological evolution. The evolutionary algorithm requires information from the objective function only, and therefore is a zero order method.

The evolutionary algorithm can be used for problems that are difficult to solve with traditional techniques, including problems that are not well defined or are difficult to model mathematically.

It can also be used when computation of the objective function is discontinuous, highly nonlinear, stochastic, or has unreliable or undefined derivatives.

This Section describes a quite general evolutionary algorithm with fitness assignment, selection, recombination and mutation. Different variants on that training operators are also explained in detail. All is implemented in the class EvolutionaryAlgorithm of the C++ library Flood [61].

Evolutionary algorithms operate on a population of individuals applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new population is created by the process of selecting individuals according to their level of fitness in the problem domain, and recombining them together using operators borrowed from natural genetics. The offspring might also undergo mutation. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation [81]. A state diagram for the training process with the evolutionary algorithm is depicted in Figure 3.5.8.

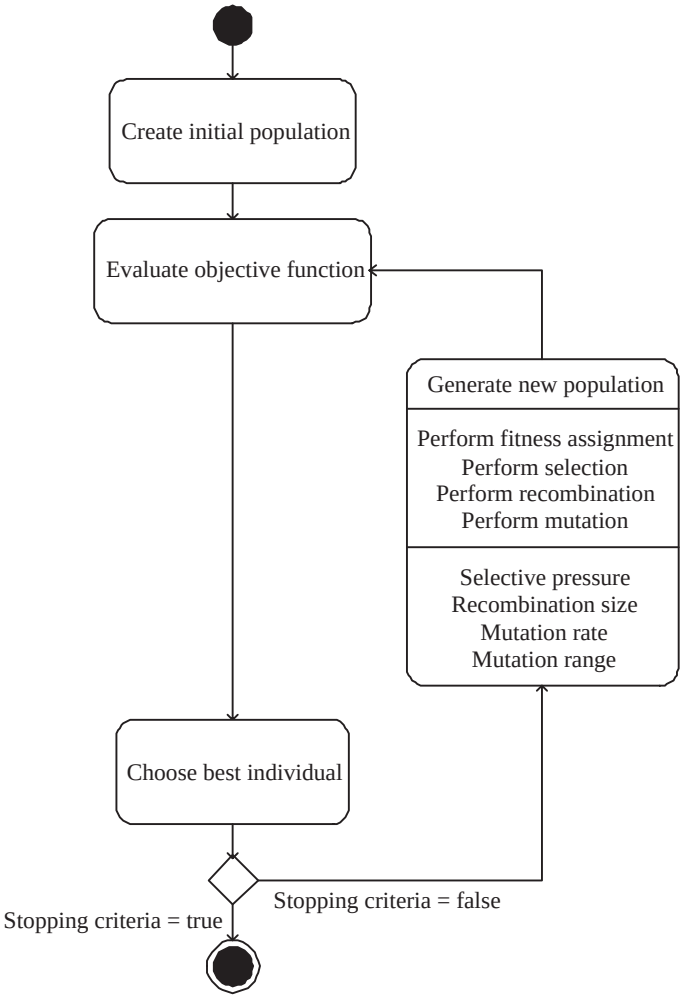


Figure 3.16: Training process with the evolutionary algorithm.

Next the training operators for the evolutionary algorithm together with their corresponding training parameters are described in detail.

Initial population

The evolutionary algorithm starts with an initial population of individuals, represented by vectors of parameters and often chosen at random

$$\mathbf{P}^{(0)} = \begin{pmatrix} \alpha_{11}^{(0)} & \dots & \alpha_{1s}^{(0)} \\ \dots & \dots & \dots \\ \alpha_{N1}^{(0)} & \dots & \alpha_{Ns}^{(0)} \end{pmatrix},$$

where $\mathbf{P}$ is called the population matrix. The number of individuals in the population N is called the population size.

The objective function is evaluated for all the individuals

$$\mathbf{f}^{(0)} = \{f_1^{(0)}(\underline{\alpha}_1^{(0)}), \dots, f_N^{(0)}(\underline{\alpha}_N^{(0)})\}, \tag{3.89}$$

being $\mathbf{f}$ the so called the evaluation vector. The individual with best evaluation $\underline{\alpha}^*$ is then chosen and stored.

Fitness assignment

If no stopping criterium is met the generation of a new population $\mathbf{P}^{(1)}$ starts by performing fitness assignment to the old population $\mathbf{P}^{(0)}$.

$$\Phi^{(0)} = \{\Phi_1^{(0)}, \dots, \Phi_N^{(0)}\}, \tag{3.90}$$

where Φ is called the fitness vector.

There are many methods of computing the fitness of the population. Proportional fitness assignment gives each individual a fitness value dependent on its actual objective function evaluation. In rank-based fitness assignment the evaluation vector is sorted. The fitness assigned to each individual depends only on its position in the individuals rank and not on the actual objective function value. Rank-based fitness assignment behaves in a more robust manner than proportional fitness assignment and, thus, is the method of choice [8] [100].

Linear ranking assigns a fitness to each individual which is linearly proportional to its rank [81]. This operator is controlled by a single parameter called selective pressure. Linear ranking allows values for the selective pressure in the interval $[1, 2]$.

Consider N the number of individuals in the population, r_i the rank of some individual i in the population, where the least fit individual has $r = 1$ and the fittest individual has $r = N$, and p the selective pressure. The fitness value for that individual is calculated as:

$$\Phi_i(r_i) = 2 - p + 2(p-1)\frac{r_i - 1}{N - 1}. \tag{3.91}$$

Selection

After fitness assignment has been performed, some individuals in the population are selected for recombination, according to their level of fitness [9] [8]. Selection determines which individuals are chosen for recombination and how many offspring each selected individual produces,

$$\mathbf{S}^{(0)} = \{S_1^{(0)}, \ldots, S_N^{(0)}\}, \tag{3.92}$$

where $\mathbf{S}$ is called the selection vector. The elements of this vector are boolean values, that is, an individual can be either selected for recombination ($true$) or not ($false$).

The simplest selection operator is roulette-wheel, also called stochastic sampling with replacement [9]. The individuals are mapped to contiguous segments of a line, such that each individual's segment is equal in size to its fitness. A random number is generated and the individual whose segment spans the random number is selected. The process is repeated until the desired number of individuals is obtained (called mating population). This technique is analogous to a roulette wheel with each slice proportional in size to the fitness, see Figure 3.17.

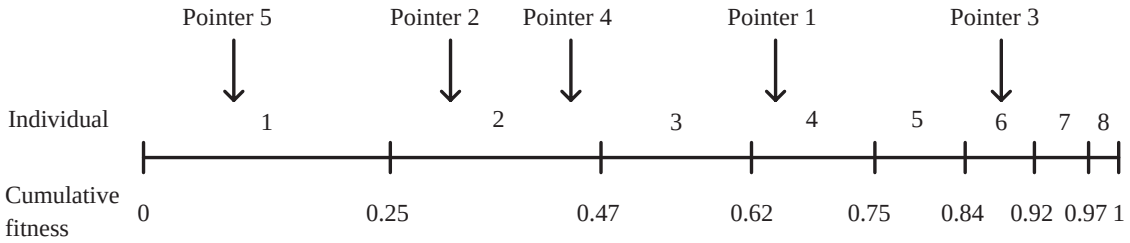


Figure 3.17: Illustration the roulette wheel selection method.

A better selection operator might be stochastic universal sampling [81]. The individuals are mapped to contiguous segments of a line, such that each individual's segment is equal in size to its fitness exactly as in roulette-wheel selection. Here equally spaced pointers are placed over the line as many as there are individuals to be selected. If the number of individuals to be selected is M, then the distance between the pointers are $1/M$ and the position of the first pointer is given by a randomly generated number in the range $[0, 1/M]$. Figure 3.18 illustrates this method.

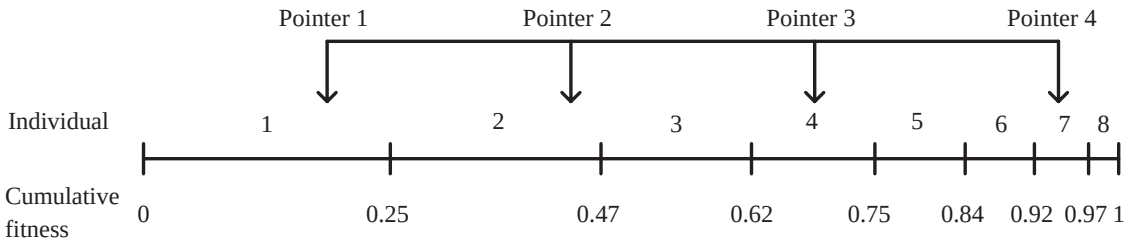


Figure 3.18: Illustration of the stochastic universal sampling selection method.

Recombination

Recombination produces a population matrix by combining the parameters of the selected individuals,

$$\mathbf{P} = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1s} \\ \dots & \dots & \dots \\ \alpha_{N1} & \dots & \alpha_{Ns} \end{pmatrix}$$

There are also many operators to perform recombination. Two of the most used are line recombination and intermediate recombination. Both line and intermediate recombination are controlled by a single parameter called recombination size, denoted d and with allowed values equal or greater than 0. In both operators, the recombination size defines the size of the area for possible offspring. A value of $d = 0$ defines the area for offspring the same size as the area spanned by the parents. Because most variables of the offspring are not generated on the border of the possible area, the area for the variables shrinks over the generations. This effect can be prevented by using a larger recombination size. A value of $d = 0.25$ ensures (statistically), that the variable area of the offspring is the same as the variable area spanned by the variables of the parents.

In line recombination the parameters of the offspring are chosen in the hyperline joining the parameters of the parents [81]. Offspring are therefore produced according to

$$\alpha_i^{(offsprint)} = a\alpha_i^{(parent1)} + (1-a)\alpha_i^{(parent2)}, \tag{3.93}$$

for $i = 1, \dots, s$ and with a chosen at random in the interval $[-d, 1+d]$.

Figure 3.19 illustrates the line recombination training operator.

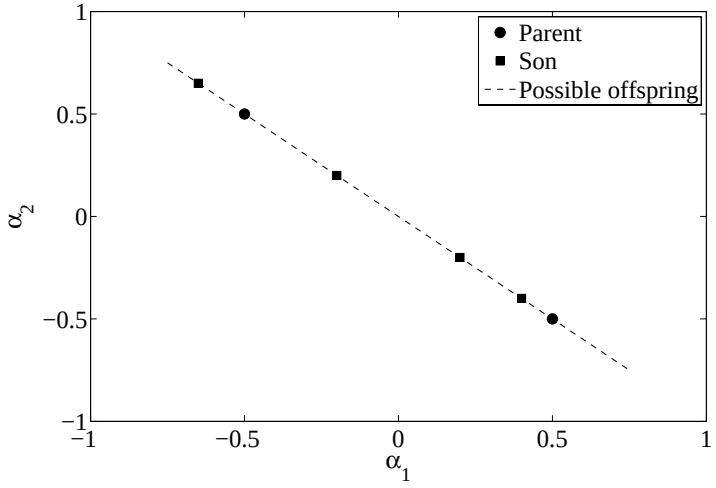


Figure 3.19: Illustration of line recombination.

Similarly, in intermediate recombination the parameters of the offspring are chosen somewhere in and around the hypercube defined by the parameters of the parents [81]. Here offspring is produced according to the rule

$$\alpha_i^{(offspring)} = a_i \alpha_i^{(parent1)} + (1 - a_i)\alpha_i^{(parent2)}, \tag{3.94}$$

for $i = 1, \ldots, s$ and with a_i chosen at random, for each i, in the interval $[-d, 1 + d]$.

Figure 3.20 is an illustration of intermediate recombination.

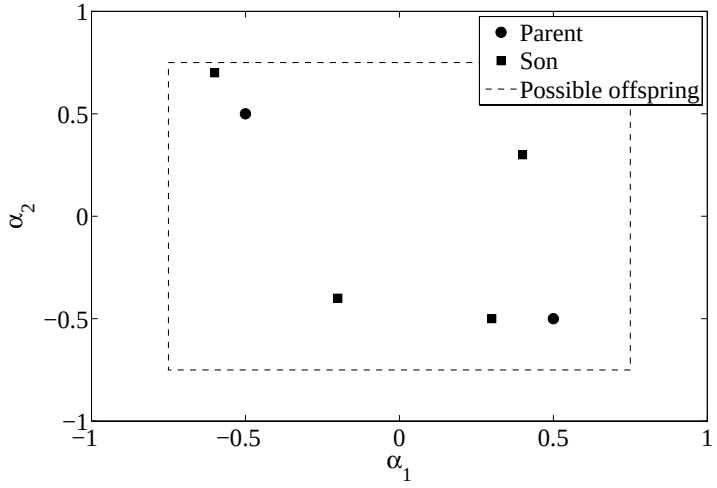


Figure 3.20: Illustration of intermediate recombination.

Mutation

Finally, some offspring undergo mutation in order to obtain the new generation,

$$\mathbf{P}^{(1)} = \begin{pmatrix} \alpha_{11}^{(1)} & \ldots & \alpha_{1s}^{(1)} \\ \ldots & \ldots & \ldots \\ \alpha_{N1}^{(1)} & \ldots & \alpha_{Ns}^{(1)} \end{pmatrix}$$

The probability of mutating a parameter is called the mutation rate and denoted p [81]. The mutation rate allows values in the interval $[0, 1]$. On the other hand, mutation is achieved by adding or subtracting a random quantity to the parameter. In this way, each parameter α_i subject to mutation is mutated to become α_i',

$$\alpha_i' = \alpha_i + \Delta\alpha_i. \tag{3.95}$$

The most common kinds of mutation procedures are uniform mutation and normal mutation. Both the uniform and normal mutation operators are controlled by a single parameter called mutation range, r, which allows values equal or greater than 0.

In uniform mutation, $\Delta\alpha_i$ is a number chosen at random in the interval $[0, r]$. Figure 3.21 illustrates the effect of uniform mutation for the case of two parameters.

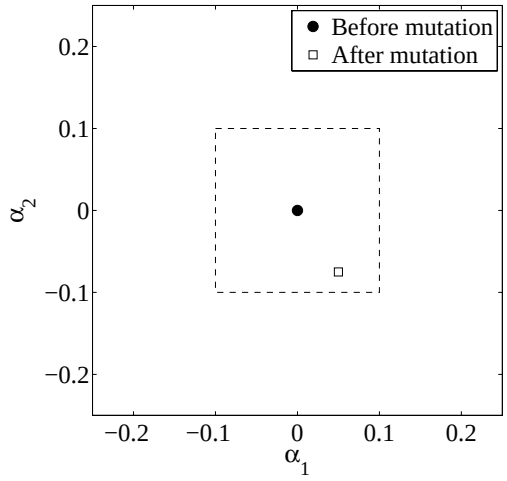


Figure 3.21: Illustration of uniform mutation.

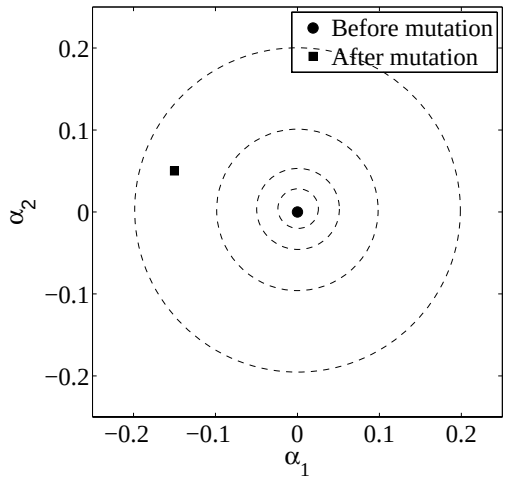


Figure 3.22: Illustration of normal mutation.

In normal mutation, $\Delta\alpha_i$ is a value obtained from a normal distribution with mean 0 and standard deviation r. Figure 3.22 illustrates the effect of normal mutation for the case of two parameters.

The whole fitness assignment, selection recombination and mutation process is repeated until a stopping criterium is satisfied.

This page is intentionally left blank.

Chapter 4

Modeling of data

Any learning task for the multilayer perceptron can be formulated as a variational problem. This statement indeed includes the traditional applications for that neural network.

In this Chapter we express the data modeling theory from a variational point of view. Actual case studies in the naval and aeronautical industries are also solved within this theory.

4.1 Problem formulation

Three main types of data modeling problems are function regression, pattern recognition and time series prediction. Some theory and typical objective functionals related to the two former is next explained.

4.1.1 Function regression

A traditional learning task for the multilayer perceptron is the function regression problem [41], which can be regarded as the problem of approximating a function from data. Here the neural network learns from knowledge represented by an input-target data set consisting of input-target examples. The targets are a specification of what the response to the inputs should be,

$$\{\mathbf{p}^{(1)}, \mathbf{t}^{(1)}\}, \{\mathbf{p}^{(2)}, \mathbf{t}^{(2)}\}, ..., \{\mathbf{p}^{(Q)}, \mathbf{t}^{(Q)}\}.$$

The basic goal in a function regression problem is to model the conditional distribution of the output variables, conditioned on the input variables [15]. This function is called the regression function.

A common feature of most input-target data sets is that the data exhibits an underlying systematic aspect, represented by some function $\mathbf{h}(\mathbf{x})$, but is corrupted with random noise. The central goal is to produce a model which exhibits good generalization, or in other words, one which makes good predictions for new data. The best generalization to new data is obtained when the mapping represents the underlying systematic aspects of the data, rather capturing the specific details (i.e. the noise contribution) of the particular input-target set. More specificially, the goal in a function regression problem for the multilayer perceptron is to obtain a function $\mathbf{y}^*(\mathbf{x})$ which approximates the regression function $\mathbf{h}(\mathbf{x})$.

One of the most common objective functionals in function regression problems is the sum squared error, described in Section 4.1.3. Also, a very common objective functional for this type of problems is the Minkowski error, which is described in Section 4.1.5.

Two frequent problems which can appear when solving a function regression problem are called underfitting and overfitting. The best generalization is achieved by using a model whose complexity is the most appropriate to produce an adequate fit of the data [27]. In this way underfitting is defined as the effect of a generalization error increasing due to a too simple model, whereas overfitting is defined as the effect of a generalization error increasing due to a too complex model.

While underfitting can be prevented by simply increasing the complexity of the neural network, it is more difficult in advance to prevent overfitting. In Section 4.1.6 we introduce the regularization theory, which is a method to prevent overfitting.

4.1.2 Pattern recognition

Another traditional learning task for the multilayer perceptron is the pattern recognition (or classification) problem [41]. The task of pattern recognition can be stated as the process whereby a received pattern, characterized by a distinct set of features, is assigned to one of a prescribed number of classes. Here the neural network learns from knowledge represented by an input-target data set consisting of input-target examples. The inputs include a set of features which characterize a pattern. The targets specify the class that each pattern belongs to,

$$\{\mathbf{p}^{(1)}, \mathbf{t}^{(1)}\}, \{\mathbf{p}^{(2)}, \mathbf{t}^{(2)}\}, ..., \{\mathbf{p}^{(Q)}, \mathbf{t}^{(Q)}\}.$$

The basic goal in a pattern recognition problem is to model the posterior probabilities of class membership, conditioned on the input variables [15]. This function is called the pattern recognition function, and it is denoted $\mathbf{h}(\mathbf{x})$.

Therefore, in order to solve a pattern recognition problem, the input space must be properly separated into regions, where each region is assigned to a class. A border between two regions is called a decision boundary. The goal in a pattern recognition problem for the multilayer perceptron is thus to obtain a function $\mathbf{y}^*(\mathbf{x})$ as an approximation of the pattern recognition function $\mathbf{h}(\mathbf{x})$.

Two of the most used objective functionals in pattern recognition problems is the sum squared error and the Minkowski error, which are described in Sections 4.1.3 and 4.1.5, respectively.

The problems of underfitting and overfitting also might occur when solving a pattern recognition problem with the sum of squares error. Underfitting is explained in terms of a too simple decision boundary which gives poor separation of the training data. On the other hand, overfitting is explained in terms of a too complex decision boundary which achieves good separation of the training data, but exhibits poor generalization.

A method for preventing underfitting and overfitting is to use a network that is just large enough to provide an adequate fit [12]. An alternative approach to obtain good generalization is by using regularization theory, which is described in Section 4.1.6.

4.1.3 The sum squared error

As it has been said, one of the most common objective functionals for the multilayer perceptron used in function regression and pattern recognition is the sum squared error (SSE). This objective functional is measured on an input target data set. The sum of the squares of the errors is used instead of the errors absolute values because this allows the objective function to be treated as a continuous differentiable function. It is written as a sum, over all the samples in the input-target

data set, of a squared error defined for each sample separately. The expression for the sum squared error is given by

$$\begin{aligned} E[\mathbf{y}(\mathbf{x};\underline{\alpha})] &= \sum_{q=1}^{Q}(E^{(q)}[\mathbf{y}(\mathbf{x};\underline{\alpha})])^2 \\ &= \sum_{q=1}^{Q}\sum_{k=1}^{m}\left(y_k(\mathbf{x}^{(q)};\underline{\alpha}) - t_k^{(q)}\right)^2, \end{aligned} \tag{4.1}$$

where Q is the number of samples in the data set, n is the number of input variables and m is the number of target variables.

There are several variant objective functionals of the sum squared error (SSE). Two of the most used are the mean squared error (MSE) and the root mean squared error (RMSE). Both objective functionals have the same properties than the sum squared error and the advantage that their value do not grow with the size of the input-target data set [15].

The expression for the mean squared error is given, in terms of the sum squared error, by

$$MSE = \frac{1}{Q}SSE, \tag{4.2}$$

and the expression for the root mean squared error is given by

$$RMSE = \sqrt{\frac{1}{Q}SSE}, \tag{4.3}$$

where, in both cases, Q is the number of samples in the data set.

The open source neural networks C++ library Flood includes the classes SumSquaredError, MeanSquaredError and RootMeanSquaredError, which represent the concepts of the sum squared error, mean squared error and root mean squared error functionals, respectively [61].

4.1.4 The normalized squared error

Another useful objective functional for data modeling problems is the normalized squared error, which takes the form

$$E[\mathbf{y}(\mathbf{x};\underline{\alpha})] = \frac{\sum_{q=1}^{Q}\|\mathbf{y}(\mathbf{x}^{(q)};\underline{\alpha}) - \mathbf{t}^{(q)}\|^2}{\sum_{q=1}^{Q}\|\mathbf{t}^{(q)} - \bar{\mathbf{t}}\|^2}. \tag{4.4}$$

The normalized squared error has the advantage that its value does not grow with the size of the input-target data set. If it has a value of unity then the neural network is predicting the data 'in the mean', while a value of zero means perfect prediction of the data [15]. As a consequence, the normalized squared error takes the same value for preprocessed data without pre and postprocessing method in the multilayer perceptron and non-preprocessed data with pre and postprocessing method in the multilayer perceptron.

Flood also includes the class NormalizedSquaredError to represent the concept of the normalized squared error [61].

4.1.5 The Minkowski error

One of the potential difficulties of the sum squared error objective functional is that it can receive a too large contribution from points which have large errors [15]. If there are long tails on the distribution then the solution can be dominated by a very small number of points which have particularly large error. In such occasions, in order to achieve good generalization, it is preferable to chose a more suitable objective functional.

We can derive more general error functions than the sum squared error for the case of a supervised learning problem. Omitting irrelevant constants, the Minkowski R-error is defined as

$$\begin{aligned} E[\mathbf{y}(\mathbf{x};\underline{\alpha})] &= \sum_{q=1}^{Q}(E^{(q)}[\mathbf{y}(\mathbf{x};\underline{\alpha})])^R \\ &= \sum_{q=1}^{Q}\sum_{k=1}^{m}|y_k(\mathbf{x}^{(q)};\underline{\alpha}) - t_k^{(q)}|^R. \end{aligned} \tag{4.5}$$

This reduces to the usual sum squared error when $R = 2$ [15].

The Minkowski error is also included in the Flood library, and it is implemented in the class MinkowskiError [61].

4.1.6 Regularization theory

A problem is called well-possed if its solution meets existence, uniqueness and stability. A solution is said to be stable when small changes in the independent variable $\mathbf{x}$ led to small changes in the dependent variable $\mathbf{y}(\mathbf{x})$. Otherwise the problem is said to be ill-possed. In this way, the function regression problem for a neural network with the sum squared error is ill-posed [24]. Indeed, the solution exists for any network architecture, but for neural networks of big complexity it might be non-unique and unstable.

In a function regression problem with the sum squared error or the Minkowski error objective functionals, the best generalization is achieved by a model whose complexity is neither too small nor too large [97]. Thus, a method for avoiding underfitting and overfitting is to use a neural network that is just large enough to provide an adequate fit. Such a neural network will not have enough power to overfit the data. Unfortunately, it is difficult to know beforehand how large a neural network should be for a specific application [38].

An alternative approach to obtain good generalization in a neural network is to control its effective complexity [96]. This can be achieved by choosing an objective functional which adds a regularization term Ω to the error functional E [27]. The objective functional then becomes

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})] = E[\mathbf{y}(\mathbf{x};\underline{\alpha})] + \nu\Omega[\mathbf{y}(\mathbf{x};\underline{\alpha})], \tag{4.6}$$

where the parameter ν is called the regularization term weight. The value of $\Omega[\mathbf{y}(\mathbf{x};\underline{\alpha})]$ typically depends on the mapping function $\mathbf{y}(\mathbf{x})$, and if the functional form $\Omega[\mathbf{y}(\mathbf{x};\underline{\alpha})]$ is chosen appropriately, it can be used to control overfitting [15].

One of the simplest forms of regularization term is called parameter decay and consists on the sum of the squares of the parameters in the neural network divided by the number of parameters [15].

$$\Omega[\mathbf{y}(\mathbf{x};\underline{\alpha})] = \frac{1}{s}\sum_{i=1}^{s}\alpha_i^2, \tag{4.7}$$

where s is the number of parameters. Adding this term to the objective function will cause the neural network to have smaller weights and biases, and this will force its response to be smoother and less likely to overfit.

The problem with regularization is that it is difficult to determine the optimum value for the term weight ν. If we make this parameter too small, we may get overfitting. If the regularization term weight is too large, the neural network will not adequately fit the data [27]. In this way, it is desirable to determine the optimal regularization parameters in an automated fashion. One approach is the Bayesian framework of David MacKay [71].

The Flood library includes the class RegularizedMeanSquaredError, which is a C++ implementation of the mean squared error with parameter decay [61].

4.1.7 Linear regression analysis

The performance of a neural network can be measured to some extent by the sum squared error on a validation data set, but it is useful to investigate the response in more detail. One option is to perform a regression analysis between the network response and the corresponding targets for an independent test set.

This analysis leads to 3 parameters. The first two, m and b, correspond to the slope and the y-intercept of the best linear regression relating targets to neural network outputs. If we had a perfect fit (outputs exactly equal to targets), the slope would be 1, and the y-intercept would be 0. The third parameter is the correlation coefficient (R-value) between the outputs and targets. If this number is equal to 1, then there is perfect correlation between targets and outputs.

The class LinearRegressionAnalysis is included in the Flood library in order to perform this validation analysis [61].

4.2 The residuary resistance of sailing yachts problem

In this Section an empirical model for the residuary resistance of sailing yachts as a function of hull geometry coefficients and the Froude number is constructed by means of a neural network. Both the data and the source code for this problem can be found within the Flood library [61]. The results from this Section were published in [80].

4.2.1 Introduction

Prediction of residuary resistance of sailing yachts at the initial design stage is of a great value for evaluating the ship's performance and for estimating the required propulsive power. Essential inputs include the basic hull dimensions and the boat velocity.

The Delft series are a semi-empirical model developed for that purpose from an extensive collection of full-scale experiments. They are expressed as a set of polynomials, and provide a prediction of the residuary resistance per unit weight of displacement, with hull geometry coefficients as variables and for discrete values of the Froude number [37]. The Delft series are widely used in the sailing yacht industry.

In this work we present a neural networks approach to residuary resistance of sailing yachts prediction. Here a multilayer perceptron has been trained with the Delf data set to provide an

estimation of the residuary resistance per unit weight of displacement as a function of hull geometry coefficients and the Froude number.

4.2.2 Experimental data

The Delft data set comprises 308 full-scale experiments, which were performed at the Delft Ship Hydromechanics Laboratory [37]. These experiments include 22 different hull forms, derived from a parent form closely related to the 'Standfast 43' designed by Frans Maas. Variations concern longitudinal position of the center of buoyancy (LCB), prismatic coefficient (C_p), length-displacement ratio ($L_{WL}/\nabla_c^{1/3}$) beam-draught ratio (B_{WL}/T_C), and length-beam ratio (L_{WL}/B_{WL}). For every hull form 14 different values for the Froude number (F_N) ranging from 0.125 to 0.450 are considered. As it has been said, the measured variable is the residuary resistance per unit weight of displacement ($1000 \cdot R_R/\triangle_c$).

Table 4.1 lists some basic statistics of these experiments. They include mean (ρ), standard deviation (σ), minimum value (min) and maximum value (max).

	LCB	C_p	$L_{WL}/\nabla_c^{1/3}$	B_{WL}/T_C	L_{WL}/B_{WL}	F_N	$(1000 \cdot R_R/\triangle_c)$
ρ	-2.38636	0.562955	4.78955	3.92455	3.20818	0.2875	10.4956
σ	1.51312	0.0220637	0.25147	0.54449	0.245434	0.100942	15.2083
min	-5	0.53	4.34	2.81	2.76	0.125	0.01
max	0	0.6	5.14	5.35	3.64	0.45	62.42

Table 4.1: Basic data set statistics in the yacht resistance problem.

4.2.3 The Delft series

The Delft series are a method for predicting the residuary resistance per unit weight of displacement of the canoe body in a sailing yacht. This model was developed from the Delf data set [37].

The Delft series are expressed as a set of 14 polynomials of the form

$$\left(\frac{1000 \cdot R_R}{\triangle_c}\right)_i = A_{i0} + A_{i1}C_p + A_{i2}C_p^2 + A_{i3}LCB + A_{i4}LCB^2 + A_{i5}\frac{B_{WL}}{T_C} + A_{i6}\frac{L_{WL}}{\nabla_c^{1/3}}. \quad (4.8)$$

for $i = 1, \ldots, 14$, and where the subindex i represents discrete values for the Froude number. The values of the coefficients A_{ij} can be found in [37].

Although the Delft series seem to track the experimental residuary resistances quite well, a standard data modeling procedure was not performed here when they were created. Indeed, the whole Delft data set was used for adjusting the polynomials (4.8), and no validation with an independent test set was done here. Also, a residuary resistance model allowing continuous values for the Froude number would be very desirable.

4.2.4 Neural networks approach

Here a multilayer perceptron is trained to learn the relationship between the input and the target variables in the Delft data set, and to make good predictions for new data.

Selection of function space

A feed-forward neural network with a sigmoid hidden layer and a linear output layer of perceptrons is used to span the function space for this problem. It must have 6 inputs (LCB, C_p, $L_{WL}/\nabla_c^{1/3}$, B_{WL}/T_C, L_{WL}/B_{WL} and F_N), and 1 output neuron ($1000 \cdot R_r/\triangle_c$).

While the numbers of inputs and output neurons are constrained by the problem, the number of neurons in the hidden layer is a design variable. In this way, and in order to draw the best network architecture, different sizes for the hidden layer are tested, and that providing the best generalization properties is adopted. In particular, the performance of three neural networks with 6, 9 and 12 hidden neurons is compared.

For that purpose, the data is divided into training, validation and testing subsets, containing 50%, 25% and 25% of the samples, respectively. More specifically, 154 samples are used here for training, 77 for validation and 77 for testing.

Table 4.2 shows the training and validation errors for the three multilayer perceptrons considered here. E_T and E_V represent the normalized squared errors made by the trained neural networks on the training and validation data sets, respectively.

Number of hidden neurons	6	9	12
E_T	0.000394	0.000223	0.000113
E_V	0.002592	0.001349	0.001571

Table 4.2: Training and validation errors in the yacht resistance problem.

As we can see, the training error decreases with the complexity of the neural network, but the validation error shows a minimum value for the multilayer perceptron with 9 hidden neurons. A possible explanation is that the lowest model complexity produces under-fitting, and the highest model complexity produces over-fitting.

In this way, the optimal number of neurons in the hidden layer turns out to be 9. This neural network can be denoted as a $6:9:1$ multilayer perceptron, and it is depicted in Figure 4.1.

The family of functions spanned by the neural network in Figure 4.1 can be denoted V and it is of dimension $s = 73$, the number of parameters.

Let denote

$$\mathbf{x} = (LCB, C_p, L_{WL}/\nabla_c^{1/3}, B_{WL}/T_C, L_{WL}/B_{WL}, F_N) \tag{4.9}$$

and

$$y = 1000 \cdot R_r/\triangle_c. \tag{4.10}$$

The elements of V are thus written

$$\begin{array}{rccc} y: & \mathbb{R}^6 & \to & \mathbb{R} \\ & \mathbf{x} & \mapsto & y(\mathbf{x}; \underline{\alpha}), \end{array}$$

where

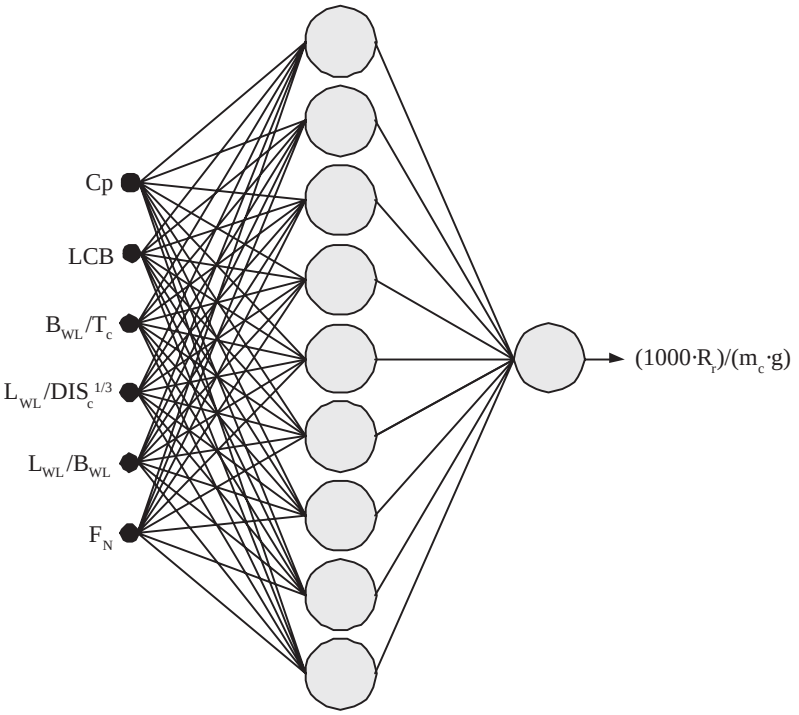


Figure 4.1: Network architecture for the yacht resistance problem.

$$y(\mathbf{x};\underline{\alpha}) = \sum_{j=0}^{9} \alpha_j^{(2)} \tanh\left(\sum_{i=0}^{6} \alpha_{ji}^{(1)} x_i\right). \tag{4.11}$$

Finally, all the biases and synaptic weights in the neural network are initialized at random.

Formulation of variational problem

The objective functional chosen for this problem is the normalized squared error between the outputs from the neural network and the target values in the Delft data set. The training data set is preprocessed with the means and the standard deviations of Table 4.1.

The variational statement of the function regression problem considered here is then to find a function $y^*(\mathbf{x};\underline{\alpha}^*) \in V$ for which the functional

$$E[y(\mathbf{x};\underline{\alpha})] \quad = \quad \frac{\sum_{q=1}^{Q} \|y(\mathbf{x}^{(q)};\underline{\alpha}) - t^{(q)}\|^2}{\sum_{q=1}^{Q} \|t^{(q)} - \bar{t}\|^2}, \tag{4.12}$$

defined on V, takes on a minimum value.

Evaluation of the objective function gradient vector is performed with the back-propagation algorithm for the normalized squared error.

Solution of reduced function optimization problem

The selected training algorithm for solving the reduced function optimization problem is a quasi-Newton method with BFGS train direction and Brent optimal train rate. Training is set to stop when the improvement between two successive epochs is less than 10^{-12}.

The evaluation and gradient norm histories are shown in Figures 4.2 and 4.3, respectively. Note that these plots have a logarithmic scale for the Y-axis.

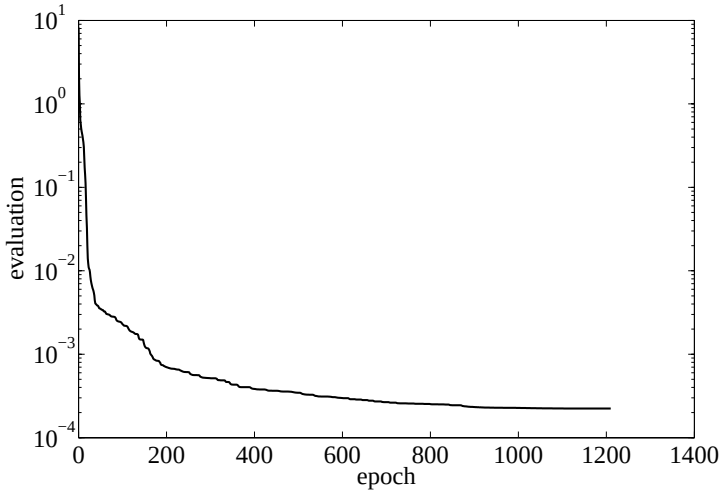


Figure 4.2: Evaluation history for the yacht resistance problem.

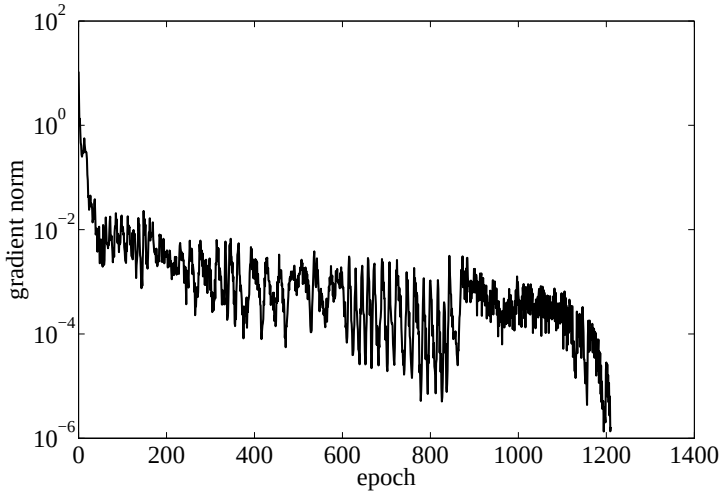


Figure 4.3: Gradient norm history for the yacht resistance problem.

Table 4.3 shows the training results for this application. Here N is the number of epochs, CPU

the training time in a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameter vector norm, $e(\underline{\alpha}^*)$ the final normalized squared error and $\|\nabla e(\underline{\alpha}^*)\|$ the final gradient norm.

N	=	1210
CPU	=	$348s$
$\|\underline{\alpha}^*\|$	=	720
$e(\underline{\alpha}^*)$	=	0.000223
$\|\nabla e(\underline{\alpha}^*)\|$	=	$1.648 \cdot 10^{-6}$

Table 4.3: Training results for the yacht resistance problem.

Once the neural network is trained, the inputs must pre-processed with the means and the standard deviations of the input data. Similarly, the outputs are be post-processed with the mean and the standard deviation of the target data.

The explicit expression for the residuary resistance model obtained by the neural network is

$$
\begin{aligned}
x_1 &= \frac{x_1 + 2.38182}{1.51322}, \\
x_2 &= \frac{x_2 - 0.564136}{0.02329}, \\
x_3 &= \frac{x_3 - 4.78864}{0.253057}, \\
x_4 &= \frac{x_4 - 3.93682}{0.548193}, \\
x_5 &= \frac{x_5 - 3.20682}{0.247998}, \\
x_6 &= \frac{x_6 - 0.2875}{0.100942}, \\
y &= 155.425 \\
&+ 63.2639 \tanh(-3.2222 + 0.0613793x_1 + 0.112065x_2 + 0.292097x_3 \\
&- 0.172921x_4 - 0.277616x_5 + 0.569819x_6) \\
&+ 91.0489 \tanh(-147.226 - 75.3342x_1 + 24.7384x_2 + 15.5625x_3 \\
&- 82.6019x_4 - 88.8575x_5 + 1.03963x_6) \\
&+ 0.00875896 \tanh(-77.0309 - 156.769x_1 - 244.11x_2 + 62.4042x_3 \\
&+ 70.2066x_4 + 12.1324x_5 - 76.0004x_6) \\
&+ 1.59825 \tanh(-2.94236 - 0.0526764x_1 - 0.21039x_2 - 0.266784x_3 \\
&+ 0.131973x_4 + 0.317116x_5 + 1.9489x_6) \\
&- 0.0124328 \tanh(-207.601 - 210.038x_1 + 99.7606x_2 + 106.485x_3 \\
&+ 252.003x_4 - 100.549x_5 - 51.3547x_6) \\
&+ 0.026265 \tanh(-17.9071 - 11.821x_1 + 5.72526x_2 - 52.2228x_3 \\
&+ 12.1563x_4 + 56.2703x_5 + 56.7649x_6) \\
&+ 0.00923066 \tanh(69.9392 - 133.216x_1 + 70.5299x_2 - 21.4377x_3 \\
&+ 47.7976x_4 + 15.1226x_5 + 100.747x_6) \\
&- 0.215311 \tanh(4.54348 - 1.11096x_1 + 0.862708x_2 + 1.61517x_3
\end{aligned}
$$

$$
\begin{aligned}
& - \quad 1.11889x_4 - 0.43838x_5 - 2.36164x_6) \\
& + \quad 0.010475 \tanh(23.4595 - 223.96x_1 - 43.2384x_2 + 13.8833x_3 \\
& + \quad 75.4947x_4 - 7.87399x_5 - 200.844x_6), \\
y^*(\mathbf{x}; \underline{\alpha}^*) & = \quad 10.4954 + 15.1605y^*(\mathbf{x}; \underline{\alpha}^*).
\end{aligned}
$$

Validation of results

A possible validation technique for the neural network model is to perform a linear regression analysis between the predicted and their corresponding experimental residuary resistance values, using and independent testing set. This analysis leads to a line $y = a + bx$ with a correlation coefficient R^2. In this way, a perfect prediction would give $a = 0$, $b = 1$ and $R^2 = 1$.

Table 4.4 shows the three parameters given by this validation analysis.

a	=	0.110
b	=	0.975
R^2	=	0.958

Table 4.4: Linear regression analysis parameters for the yacht resistance problem.

Figure 4.4 illustrates a graphical output provided by this validation analysis. The predicted residuary resistances are plotted versus the experimental ones as open circles. The solid line indicates the best linear fit. The dashed line with $R^2 = 1$ would indicate perfect fit.

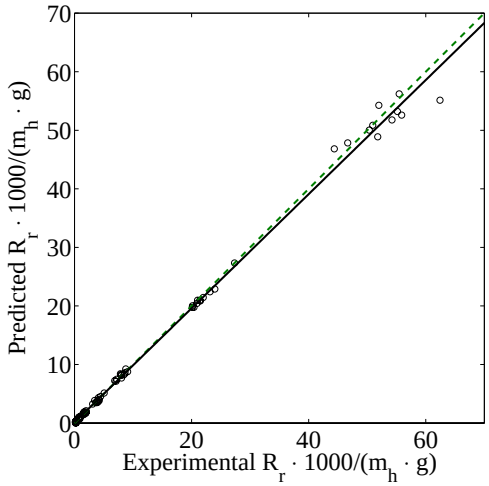


Figure 4.4: Linear regression analysis plot for the yacht resistance problem.

From Table 4.4 and Figure 4.4 we can see that the neural network is predicting very well the entire range of residuary resistance data. Indeed, the a, b and R^2 values are very close to 0, 1 and 1, respectively.

4.2.5 Conclusions

A neural network model has been developed for estimating the residuary resistance of sailing yachts, and it was found that the quality of the prediction on a testing data set is very satisfactory over the entire range of data. Moreover, the approach presented here allows continuous values for the Froude number, while the Delf series do not.

4.3 The airfoil self-noise problem

A neural network model for airfoil self-noise prediction has been created using a noise spectrum database collected from a family of NACA 0012 airfoils. This application is included within the Flood library [61] and it has been published as [57].

4.3.1 Introduction

The noise generated by an aircraft is an efficiency and environmental matter for the aerospace industry. NASA have issued a mandate to reduce the external noise generated by the whole airframe of an aircraft by 10 decibels (dB) in the near term future [60].

An important component of the total airframe noise is the airfoil self-noise, which is due to the interaction between an airfoil blade and the turbulence produce in its own boundary layer and near wake.

There have been many attempts to predict the noise of an airfoil. Howe [50] reviewed the various methodologies and grouped the different approaches into three groups: (i) theories based upon the Lighthill acoustic analogy [59], (ii) theories based on the solution of special problems approximated by the linearized hydrodynamics equations [2], and (iii) semi-empirical models [16].

The model of Brooks, Pope and Marcolini [16], referred here as the BPM model, is a semi-empirical approach. It is formulated from an extensive set of acoustic wind tunnel tests performed by NASA, upon different chord length NACA 0012 airfoil sections. The BPM model is used to predict the self-generated noise of an airfoil blade encountering smooth flow as a function of the angle of attack, the free-stream velocity and the geometric parameters of the airfoil. Formulated in 1989, the BPM model is still utilized nowadays.

However, it can be shown that the BPM model does not predict accurately the SPL variable. For example, it substantially underestimates the airfoil self-noise at low values. In this regard, a new, more accurate, model to predict the SPL generated by an airfoil is necessary.

Technologies based on neural networks are currently being developed which may assist in addressing a wide range of problems in aeronautics. They have already been successfully applied to a range of aeronautical problems in fault diagnostics, modeling and simulation and control systems [32].

In this work a neural networks approach to airfoil self-noise prediction is presented. After reviewing the prediction quality of all the configurations tested, it is shown that the neural network model provides good results for an independent testing data set across the entire range of sound pressure level data.

4.3.2 Experimental data

The self-noise data set used in this work was processed by NASA in 1989 [16], and so it is referred here to as the NASA data set. It was obtained from a series of aerodynamic and acoustic tests of two and three-dimensional airfoil blade sections conducted in an anechoic wind tunnel. A complete description of these experiments is reported in [16].

The NASA data set comprises different size NACA 0012 airfoils at various wind tunnel speeds and angles of attack. The span of the airfoil and the observer position were the same in all of the experiments.

The aim of the acoustic measurements was to determine spectra for self-noise from airfoils encountering smooth airflow. These spectra are affected by the presence of extraneous contributions, and some processing of the data was needed. The results were presented as 1/3-octave spectra.

In that way, the NASA data set contains 1503 entries and consists of the following variables:

1. Frequency, $f\,[Hz]$.
2. Angle of attack, $\alpha\,[^\circ]$.
3. Chord length, $c\,[m]$.
4. Free-stream velocity, $U\,[ms^{-1}]$.
5. Suction side displacement thickness, $s\,[m]$.
6. Scaled sound pressure level, $SPL_{1/3}\,[dB]$.

Here the suction side displacement thickness was determined using an expression derived from boundary layer experimental data from [16].

Table 4.5 depicts some basic statistics of the NASA data set. Here ρ is the mean, σ the standard deviation, min the minimum and max the maximum.

Variable	$f\,[Hz]$	$\alpha\,[^\circ]$	$c\,[m]$	$U\,[ms^{-1}]$	$s\,[m]$	$SPL_{1/3}\,[dB]$
ρ	2886.38	6.7823	0.136548	50.8607	0.0111399	124.836
σ	3152.57	5.91813	0.0935407	15.5728	0.0131502	6.89866
min	200	0	0.0254	31.7	0.000400682	103.38
max	20000	22.2	0.3048	71.3	0.0584113	140.987

Table 4.5: Basic input-target data set statistics in the airfoil noise problem.

4.3.3 The Brooks-Pope-Marcolini (BPM) model

The BPM model is a semi-empirical approach based on previous theoretical studies and calibrated using the NASA data set [16].

The airfoil self-noise is the total noise produced when an airfoil encounters smooth non-turbulent inflow. In this way Brooks et al. identified and modeled five self-noise mechanisms due to specific boundary-layer phenomena [16]: (i) boundary-layer turbulence passing the training edge, (ii) separated-boundary-layer and stalled-airfoil flow, (iii) vortex shedding due to laminar-boundary-layer instabilities, (iv) vortex shedding from blunt trailing edges, and (v) turbulent vortex flow existing near the tips of lifting blades. This five noise generation mechanisms are illustrated in Figure 4.5.

The total sound pressure level is determined by summing up the contributions from each of the different phenomena enumerated above. The particular noise mechanisms which contribute to the total sound pressure level are dependent on the flow conditions upon the airfoil. In the Brooks model there are two distinct boundary layer conditions: tripped and untrippred (or natural). Depending on which is present the total self-noise expression has a different form.

For the tripped boundary layer condition the laminar boundary layer noise mechanism cannot occur, as tripping the boundary layer causes it to become turbulent. Therefore its contribution to the total sound pressure level is ignored, and only the trailing edge noise is evaluated. In the BPM model the trailing edge noise consists of noise generated by separation (SPL_α), and pressure and suction side effects (SPL_P and SPL_S). The total self-noise for the tripped boundary layer condition is as follows,

$$SPL_{Total} = 10 \log \left(10^{SPL_P/10} + 10^{SPL_S/10} + 10^{SPL_\alpha/10}\right). \tag{4.13}$$

Conversely, if the boundary layer is untripped, then the laminar boundary layer vortex shedding noise mechanism must also be considered. This contribution is termed SPL_{LBL-VS}. The total sound pressure level for the untripped boundary layer condition is then given by the following equation,

$$SPL_{Total} = 10 \log \left(10^{SPL_P/10} + 10^{SPL_S/10} + 10^{SPL_\alpha/10} + 10^{SPL_{LBL-VS}/10}\right). \tag{4.14}$$

In this case study we will only deal with the tripped boundary layer condition.

4.3.4 Neural networks approach

In this Section, a noise prediction model for a turbulent boundary layer noise mechanism has been created using a multilayer perceptron trained to learn the underlying aspects of the NASA data set.

Selection of function space

Here a multilayer perceptron with a sigmoid hidden layer and a linear output layer is used to predict airfoil self-noise ($SPL_{1/3}$) as a function of frequency (f), angle of attack (α), chord length (c), suction side displacement thickness (s) and free-stream speed (U). Therefore, this neural network must have 5 inputs, f, α, c, s and U, and 1 output neuron, $SPL_{1/3}$.

The number of neurons in the hidden layer will define the degree of complexity of the airfoil self-noise model, and it needs to be chosen carefully. Indeed, two frequent problems which can appear when solving a data modeling problem are called under-fitting and over-fitting. The best generalization is achieved by using a model whose complexity is the most appropriate to produce an adequate fit of the data [27]. In this way under-fitting is defined as the effect of a generalization error increasing due to a too simple model, whereas over-fitting is defined as the effect of a generalization error increasing due to a too complex model.

In this way, and in order to draw a correct network architecture for this problem, the NASA data set was first split into three, training, validation and testing, subsets. They represent 50%, 25% and 25% of the full data set, respectively. Therefore the training data set contains 752 samples, the validation data set contains 376 samples and the testing data set contains 375 samples.

Three different neural networks with 6, 9 and 12 neurons in the hidden layer are then trained with the normalized squared error objective functional and the quasi-Newton method training algorithm. The generalization performance of each single neural network was then assessed by measuring the error between the outputs from the trained neural networks and the target values in the validation data set. This error can indicate if the current model is under or over-fitting the data. Table 4.6 shows the training and the validation errors for the three different network architectures considered. The training error decreases with the complexity of the neural network,

but the validation error shows a minimum value for the multilayer perceptron with 9 hidden neurons. Indeed, the lowest model complexity seems to produce under-fitting, while the highest model complexity seems to produce over-fitting.

Number of hidden neurons	6	9	12
Training error	0.122405	0.082553	0.060767
Validation error	0.131768	0.116831	0.129884

Table 4.6: Training and validation errors in the airfoil noise problem.

From these results follows that the optimal network architecture for this problem is that including 9 hidden neurons. This neural network can be denoted as a $5:9:1$ multilayer perceptron, and it is depicted in Figure 4.6.

This multilayer perceptron spans a function space V of dimension $s = 64$. If we denote $\mathbf{x} = (f, \alpha, c, U, s)$ and $y = SPL_{1/3}$ the functions belonging to that family are of the form

$$\begin{array}{rccc} y: & \mathbb{R}^5 & \to & \mathbb{R} \\ & \mathbf{x} & \mapsto & y(\mathbf{x}; \underline{\alpha}), \end{array}$$

where

$$y(\mathbf{x}; \underline{\alpha}) = \sum_{j=0}^{9} \alpha_{kj}^{(2)} \tanh \left(\sum_{i=0}^{5} \alpha_{ji}^{(1)} x_i \right). \tag{4.15}$$

Here the parameter vector $\underline{\alpha}$ is randomly initialized.

Formulation of variational problem

First, the training data set is preprocessed with the means and standard deviations of the NASA data set in Table 4.5.

A statistical analysis of the NASA data set did not show a significant presence of outliers. On the other hand, the optimal network architecture for the validation data has been selected, which means that the multilayer perceptron will not over-fit the data. In this way the simple normalized squared error is used instead the more complex Minkowski error or Bayesian regularization.

Therefore, the variational statement of this problem is to find a function $y^*(\mathbf{x}; \underline{\alpha}^*) \in V$ for which the functional

$$E[y(\mathbf{x}; \underline{\alpha})] = \frac{\sum_{q=1}^{Q} \| y(\mathbf{x}^{(q)}; \underline{\alpha}) - t^{(q)} \|^2}{\sum_{q=1}^{Q} \| t^{(q)} - \bar{t} \|^2}, \tag{4.16}$$

defined on V, takes on a minimum value.

The objective function gradient $\nabla e(\underline{\alpha})$ is computed with a back-propagation algorithm for the normalized squared error.

Solution of reduced function optimization problem

The training algorithm used to optimize the associated objective function is a quasi-Newton method with BFGS train direction and Brent train rate.

The training process is set to stop when improvement between two successive epochs is less than 10^{-12}. Figures 4.7 and 4.7 show the history of this process for the evaluation and the gradient norm, respectively. Note that these plots have a logarithmic scale for the Y-axis. The evaluation history shows a significant reduction of this variable during the first few epochs and a posterior stabilization. On the other hand, during the training process the gradient norm decreases in several orders of magnitude. These two signs demonstrate that the training algorithm has certainly reached a minimum.

Some final results from training are listed in Table 4.7. Here N is the number of epochs, CPU the computing time in a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameters norm, $e(\underline{\alpha}^*)$ the final normalized squared error and $\|\nabla e(\underline{\alpha}^*)\|$ its gradient norm.

N	=	2031
CPU	=	$2218s$
$\|\underline{\alpha}^*\|$	=	11257.4
$e(\underline{\alpha}^*)$	=	0.082553
$\|\nabla e(\underline{\alpha}^*)\|$	=	0.011912

Table 4.7: Training results for the airfoil self-noise problem.

The explicit expression of the neural network model for airfoil self-noise prediction is

$$
\begin{aligned}
x_1 &= \frac{x_1 - 2886.38}{3152.57}, \\
x_2 &= \frac{x_2 - 6.7823}{5.91813}, \\
x_3 &= \frac{x_3 - 0.136548}{0.0935407}, \\
x_4 &= \frac{x_4 - 50.8607}{15.5728}, \\
x_5 &= \frac{x_5 - 0.0111399}{0.0131502}, \\
y^*(\mathbf{x}, \underline{\alpha}^*) &= -3.26798 \\
&- 1.97066 \tanh(2.28999 + 0.0677432x_1 + 0.198962x_2 - 0.224465x_3 + 0.0961816x_4 + 2.7212x_5) \\
&+ 4.69623 \tanh(3.42765 + 2.34839x_1 + 0.635714x_2 + 0.0415776x_3 - 0.0832574x_4 + 2.30865x_5) \\
&+ 0.199113 \tanh(-4010.32 - 3442.27x_1 - 2549.77x_2 + 252.083x_3 - 281.831x_4 + 2783.93x_5) \\
&+ 1921.5 \tanh(0.444241 + 0.148645x_1 - 0.0397378x_2 - 0.371608x_3 + 0.102558x_4 + 0.0102109x_5) \\
&- 0.433932 \tanh(0.618045 + 1.44893x_1 + 3.03601x_2 + 2.75553x_3 - 0.669881x_4 - 1.16721x_5) \\
&- 0.313583 \tanh(3824 + 3180.03x_1 + 2505.72x_2 + 6753.55x_3 - 303.828x_4 - 561.2x_5) \\
&+ 1.71381 \tanh(6.81775 + 8.18818x_1 - 0.222292x_2 + 0.430508x_3 - 0.152801x_4 + 0.801288x_5) \\
&- 3.91 \tanh(2.20453 + 2.68659x_1 + 0.96715x_2 + 0.0871504x_3 - 0.102282x_4 + 0.00203128x_5) \\
&- 1917.76 \tanh(0.443727 + 0.149241x_1 - 0.0404185x_2 - 0.372191x_3 + 0.102622x_4 + 0.0107115x_5), \\
y^*(\mathbf{x}, \underline{\alpha}^*) &= 124.836 + 6.89866y^*(\mathbf{x}, \underline{\alpha}^*).
\end{aligned}
$$

Validation of results

The accuracy of the neural network model over the experimental configurations in the testing data set is measured by performing a linear regression analysis between the outputs from the neural network and the corresponding targets.

The values of the linear regression parameters here are listed in Table 4.8. Here a is the x-intercept, b is the slope and R^2 is the correlation coefficient.

a	=	12.5411
b	=	0.900159
R^2	=	0.894

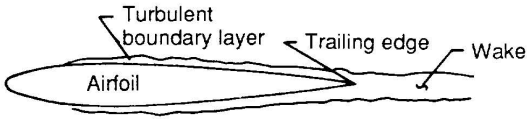
Table 4.8: Linear regression analysis parameters for the airfoil self noise problem.

Figure 4.9 is a plot of this validation analysis. The broken line represents the optimal model, in which the prediction would perfectly match the experimental data. From this figure we can see that the neural network model provides a good fit across all of the equipment configurations.

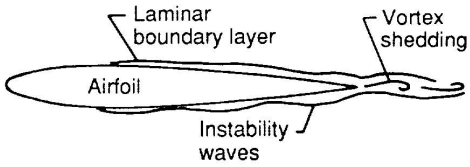
4.3.5 Conclusions

An overall prediction method based on neural networks has been developed. The approach is based on data from a series of aerodynamic and acoustic tests of isolated airfoil sections.

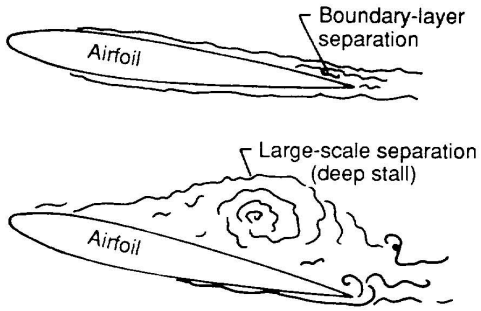
The results of the neural network model were checked against an independent testing data set, and it was found that the quality of the prediction was satisfactory over the entire range of sound pressure level data.



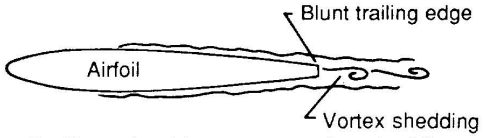
Turbulent-boundary-layer—trailing-edge noise



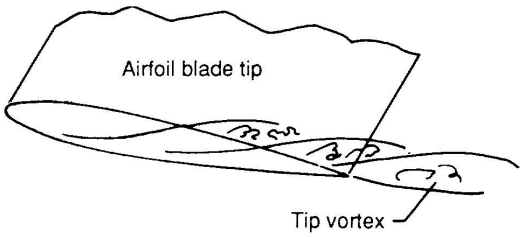
Laminar-boundary-layer—vortex-shedding noise



Separation-stall noise



Trailing-edge-bluntness—vortex-shedding noise



Tip vortex formation noise

Figure 4.5: Airfoil self-noise mechanisms.

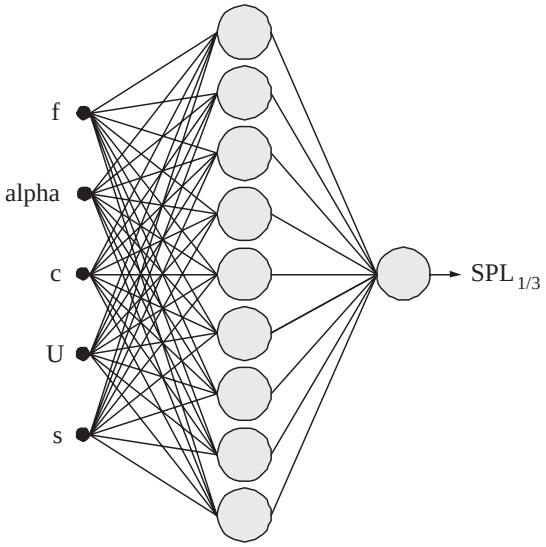


Figure 4.6: Optimal network architecture for the airfoil self-noise problem.

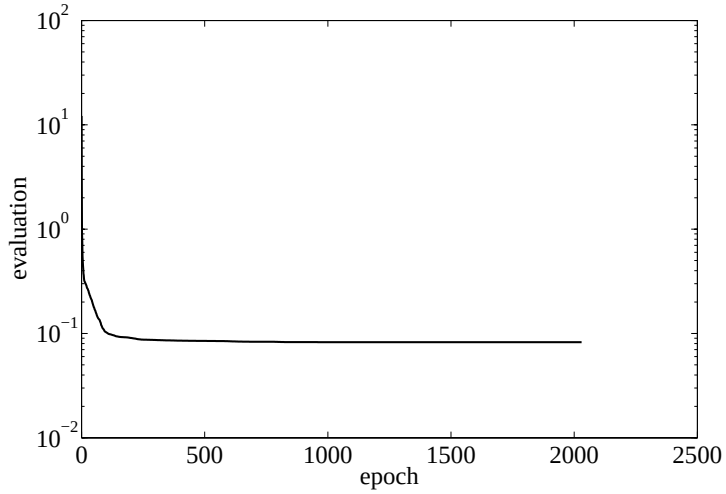


Figure 4.7: Evaluation history for the airfoil self-noise problem problem.

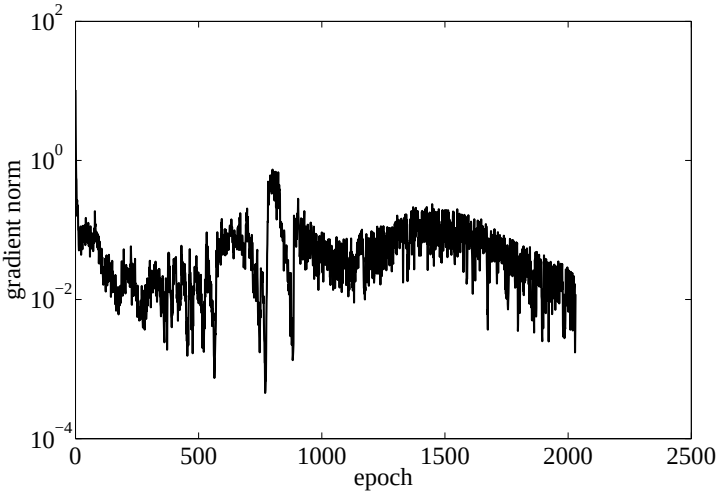


Figure 4.8: Gradient norm history for the airfoil self-noise problem.

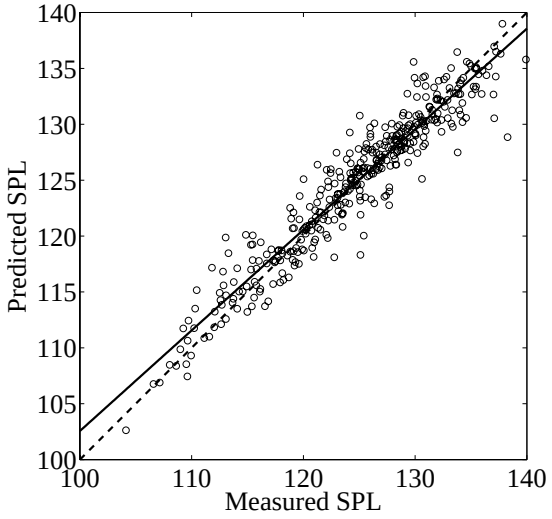


Figure 4.9: Linear regression analysis in the airfoil noise problem.

Chapter 5

Classical problems in the calculus of variations

A variational formulation for the multilayer perceptron provides a direct method for variational problems. Within this formulation, the solution approach consists of three steps, selection of function space, formulation of variational problem and solution of reduced function optimization problem.

For the purpose of validating this numerical method, neural computing is used in this Chapter to solve some classical problems in the calculus of variations. The neural network answers are compared to the analytical ones, with encouraging results. It is very important to understand that no input-target data is used in any of these examples. Instead, the neural networks learn here from mathematical models.

5.1 The geodesic problem

The geodesic problem for the multilayer perceptron is a variational problem with one input and one output variables, two boundary conditions and where evaluation of the objective functional is obtained by integrating a function. All the source code used to solve this problem is included in the C++ library Flood [61].

5.1.1 Problem statement

The geodesic problem on the plane can be stated as:

Given two points $A = (x_a, y_a)$ and $B = (x_b, y_b)$ in a plane, find the shortest path between A and B.

Figure 5.1 depicts graphically the formulation for this case study.

The arc length between point A and point B of a curve $y(x)$ is given by the functional

$$L[y(x)] = \int_A^B ds$$

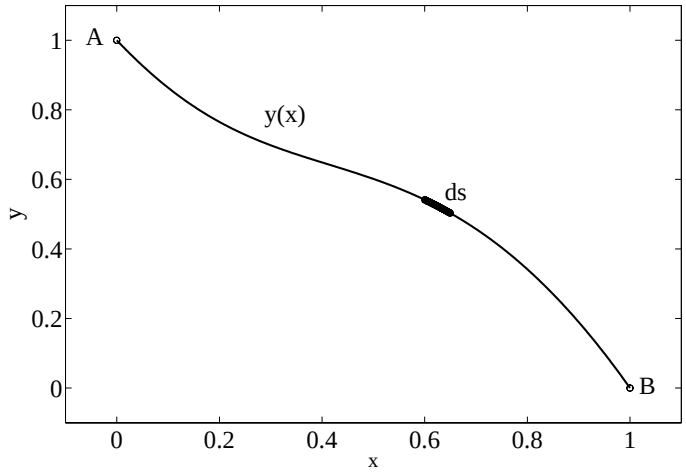


Figure 5.1: The geodesic problem statement.

$$\begin{aligned} &= \int_{x_a}^{x_b} \sqrt{dx^2 + dy^2} \\ &= \int_{x_a}^{x_b} \sqrt{1 + [y'(x)]^2} dx. \end{aligned} \tag{5.1}$$

The analytical solution to the geodesic problem in the plane is obviously a straight line. For the particular case when $A = (1, 0)$ and $B = (0, 1)$, the Euler-Lagrange equation provides the following function as the minimal value for the functional in Equation (5.1)

$$y^*(x) = 1 - x, \tag{5.2}$$

which gives $L[y^*(x)] = 1.414214$.

5.1.2 Selection of function space

The neural network chosen to span a function space for the geodesic problem is a multilayer perceptron with one input, x, a sigmoid hidden layer of size three and one linear output neuron, y. Such a computational device is drawn in Figure 5.2.

The function space spanned by this neural network is denoted V and it is of dimension $s = 10$, the number of parameters. Elements of V are written

$$\begin{aligned} y: \quad \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto y(x; \underline{\alpha}), \end{aligned}$$

with

$$y(x; \underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} x\right). \tag{5.3}$$

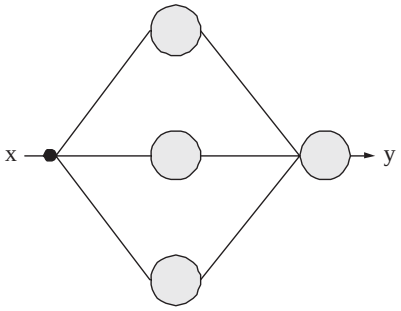


Figure 5.2: Network architecture for the geodesic problem.

The geodesic problem stated in this Section is required to hold $y(0) = 1$ and $y(1) = 0$, the boundary conditions. Therefore this multilayer perceptron must be extended with a particular and an independent solution terms. A suitable set here is

$$\varphi_0(x) = (1 - x), \tag{5.4}$$
$$\varphi_1(x) = x(x - 1), \tag{5.5}$$

respectively. They give

$$y(x; \underline{\alpha}) = (1 - x) + x(x - 1)y(x; \underline{\alpha}). \tag{5.6}$$

Note that all the functions in Equation (5.6) indeed satisfy $y(0) = 1$ and $y(1) = 0$.

On the other hand, all the biases and synaptic weights in the neural network are initialized here at random. The resulting initial guess is depicted in Figure 5.3.

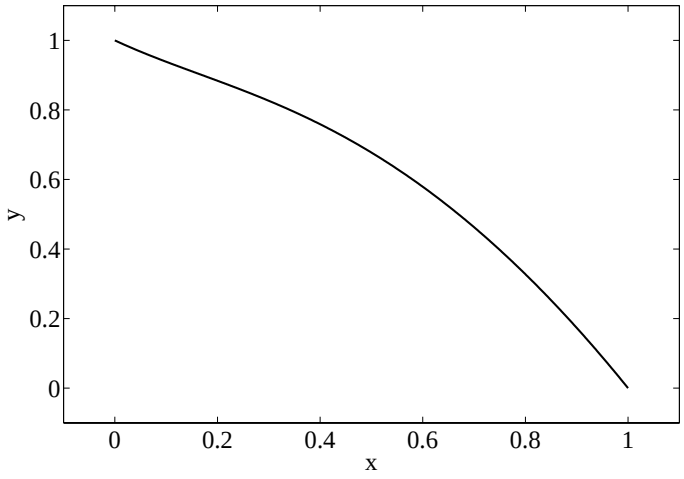


Figure 5.3: Initial guess for the geodesic problem.

5.1.3 Formulation of variational problem

From Equation (5.1), the variational statement of the geodesic problem for the multilayer perceptron is to find a function $y^*(x;\underline{\alpha}^*) \in V$ for which the functional

$$L[y(x;\underline{\alpha})] = \int_{x_a}^{x_b} \sqrt{1 + [y'(x;\underline{\alpha})]^2} dx, \tag{5.7}$$

defined on V, takes on a minimum value.

The derivative of the output with respect to the input, $y'(x;\underline{\alpha})$, is computed with the back-propagation algorithm for the Jacobian matrix.

On the other hand, calculating the arc-length of some function $y(x;\underline{\alpha}) \in V$ requires to integrate a function. Here we use an ordinary differential equation approach for that purpose. In this way, the value of the integral in Equation (5.7) is equivalent to obtain the value $y(b)$ which is the solution to

$$\frac{dy}{dx} = \sqrt{1 + [y'(x)]^2}, \tag{5.8}$$

$$y(a) = y_a. \tag{5.9}$$

This Cauchy problem can be solved with an adaptive stepsize algorithm. Here we use the Runge-Kutta-Fehlberg method with a tolerance of 10^{-15}.

Finally, a back-propagation algorithm for the objective function gradient, $\nabla f(\underline{\alpha})$, is not possible to be derived here, since the target outputs from the neural network are not known. Instead, the central differences method for numerical differentiation is to be used, with $\epsilon = 10^{-6}$.

5.1.4 Solution of reduced function optimization problem

A quasi-Newton method is the training algorithm used to solve the reduced function optimization problem in this application. The particular train direction chosen is that by the BFGS method and the train rate is computed with the Brent's method.

The training algorithm is set to stop when it is not able to reduce the evaluation value of the objective function anymore. The arc-length of the randomly initialized neural network is 1.44235. The quasi-Newton method requires 73 epochs to converge. The final value of the arc-length is 1.41421. On the other side the gradient norm decreases from 0.0588223 to $7.61254 \cdot 10^{-7}$. Figure 5.4 shows the evaluation value of the neural network as a function of the training epoch, while Figure 5.5 depicts the gradient norm training history with a logarithmic scale for the y-axis. There is a significant decrease in both the evaluation value and the gradient norm.

Some numbers resulted from this training process are listed in Table 5.1. These include the number of epochs N, the number of calls to the objective function M, the CPU time in a laptop AMD 3000 CPU, the final parameter vector norm $\|\underline{\alpha}^*\|$, the final objective value $l(\underline{\alpha}^*)$ and the final gradient norm $\|\nabla l(\underline{\alpha}^*)\|$. The final arc-length here yields a percentage error of the neural network results with respect to the analytical solution of about $-2.519 \cdot 10^{-4}\%$.

It is also possible to write an explicit expression of the function addressed by the multilayer perceptron. This is given by

$$\begin{aligned} y^*(\mathbf{x};\underline{\alpha}^*) &= (1-x) + x(x-1)[0.589469 \\ &- 1.20624 \tanh(0.29317 + 0.00800689x) \end{aligned}$$

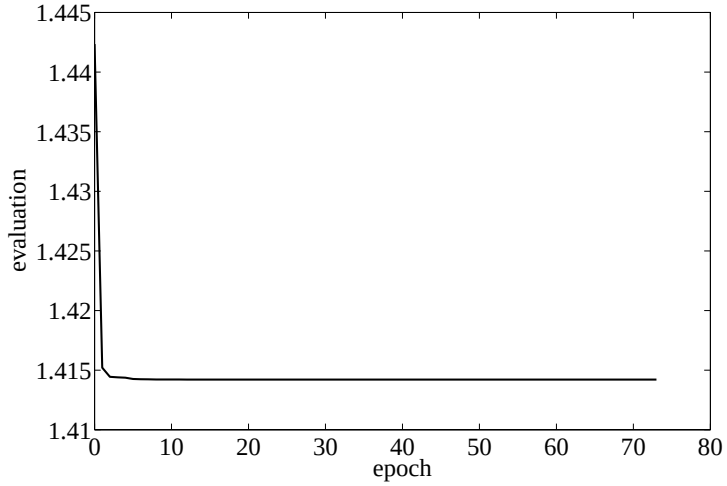


Figure 5.4: Evaluation history for the geodesic problem.

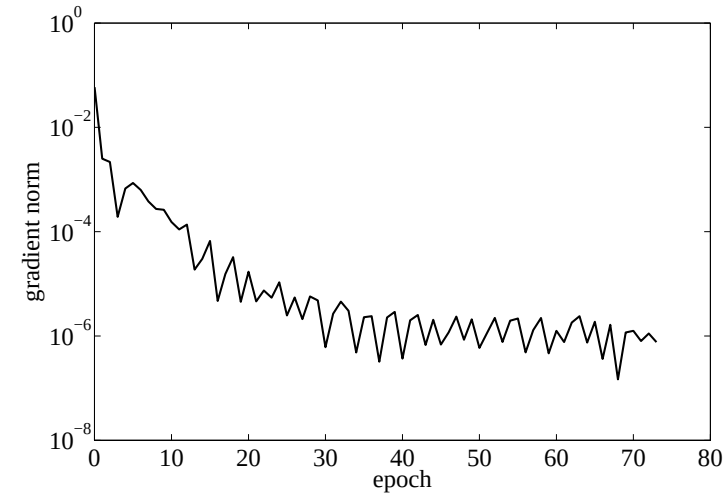


Figure 5.5: Gradient norm history for the geodesic problem.

$$\begin{aligned} &- \quad 0.955752 \tanh\left(2.27028 - 0.175609x\right) \\ &- \quad 0.691202 \tanh\left(-3.59022 - 0.405461x\right)], \end{aligned} \tag{5.10}$$

for $x \in [0, 1]$.

To conclude, the final shape of the neural network is shown in Figure 5.6. As we can see, this is very much a straight line joining the two points A and B, which is the analytical solution for the geodesic problem.

N	=	73
M	=	4260
CPU	=	$62s$
$\|\underline{\alpha}^*\|$	=	1.46693
$l(\underline{\alpha}^*)$	=	1.41421
$\|\nabla l(\underline{\alpha}^*)\|$	=	$7.61254 \cdot 10^{-7}$

Table 5.1: Training results for the geodesic problem.

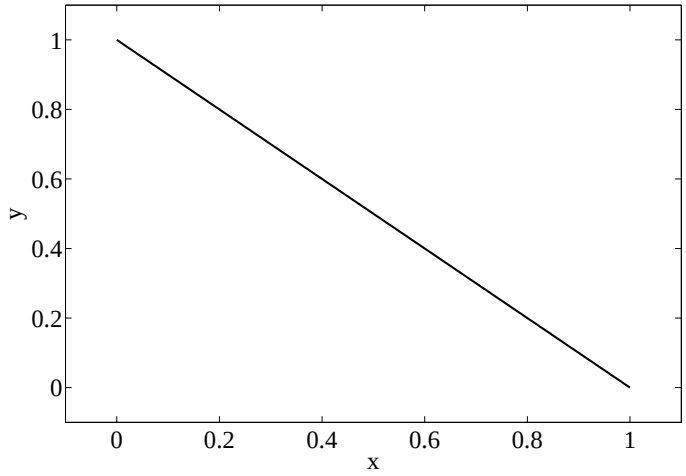


Figure 5.6: Neural network results for the geodesic problem.

5.2 The brachistochrone problem

The brachistochrone problem for the multilayer perceptron can be stated as a variational problem with one input and one output variables, two boundary conditions and an objective functional defined by an improper integral. A C++ software implementation of this problem can also be found within Flood [61]. Similar results to those included here are published in [66].

5.2.1 Problem statement

The original statement of the brachistochrone problem is:

Given two points $A = (x_a, y_a)$ and $B = (x_b, y_b)$ in a vertical plane, what is the curve traced out by a particle acted on only by gravity, which starts at A and reaches B in the shortest time?

In this example we employ a cartesian representation, i.e., the brachistochrone is chosen to be of the form $y = y(x)$. Nevertheless, other representations would be possible, such as a parametric representation, in which the brachistochrone would take the form $(x, y) = (x(t), y(t))$, see Figure 5.2.1.

The time for a particle to travel from point A to point B within a curve $y(x)$ is given by the

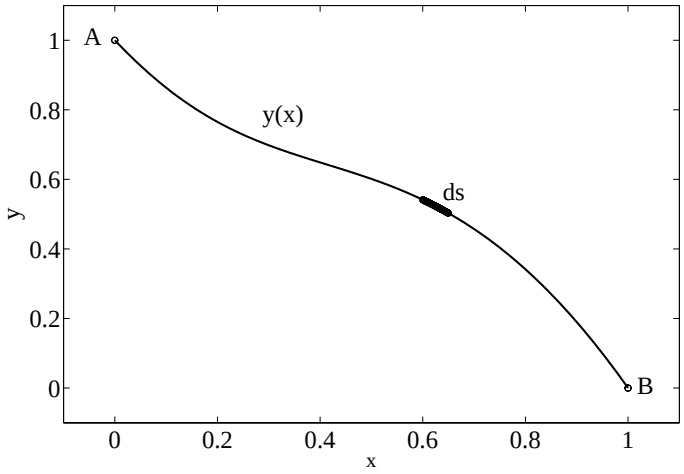


Figure 5.7: The brachistochrone problem statement.

functional

$$T[y(x)] = \int_A^B \frac{ds}{v}, \tag{5.11}$$

where s is the arc length and v is the speed. The arc length element here can be written

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2} \\ &= \sqrt{1 + [y'(x)]^2}dx, \end{aligned} \tag{5.12}$$

On the other hand, and in order to obtain an expression for the speed, consider the law of conservation of energy,

$$mgy_a = mgy(x) + \frac{1}{2}mv^2, \tag{5.13}$$

where $g = 9.81$ is the gravitational acceleration. This gives

$$v = \sqrt{2g(y_a - y(x))}. \tag{5.14}$$

Plugging (5.12) and (5.14) into (5.11) produces the final expression for the descent time

$$T[y(x)] = \frac{1}{\sqrt{2g}} \int_{x_a}^{x_b} \sqrt{\frac{1 + [y'(x)]^2}{y_a - y(x)}}dx. \tag{5.15}$$

The analytical solution to the brachistochrone problem is a segment of cycloid [99]. Taking $A = (0, 1)$ and $B = (1, 0)$, the optimal function is given in parametric equations by

$$x^*(\theta) = 0.583(\theta - \sin\theta), \tag{5.16}$$
$$y^*(\theta) = 1 - 0.583(1 - \cos\theta), \tag{5.17}$$

for $\theta \in [0, 2.412]$. Equations (5.16) and (5.17) yield an optimum descent time $T[(x, y)^*(\theta)] = 0.577$. The analytical solution to this particular problem can be found in Section 2.2.

5.2.2 Selection of function space

Here a multilayer perceptron with a sigmoid hidden layer and a linear output layer is used to represent the descent curve $y(x)$. As it is to be described in cartesian coordinates, the neural network must have one input, x, and one output neuron, y. The size of the hidden layer is a design variable. However, this is not a very critical issue here, since this problem is well-posed, and we use three neurons in the hidden layer. Figure 5.8 is a graphical representation of this network architecture.

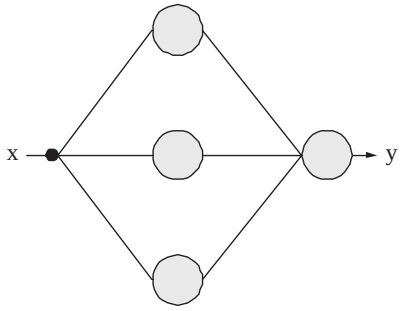


Figure 5.8: Network architecture for the brachistochrone problem.

Such a multilayer perceptron spans a family V of parameterized functions $y(x; \underline{\alpha})$ of dimension $s = 10$, which is the number of parameters in the neural network. Elements of V are of the form

$$\begin{aligned} y: \quad \mathbb{R} &\to \mathbb{R} \\ x &\mapsto y(x; \underline{\alpha}), \end{aligned}$$

where

$$y(x; \underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} x\right). \tag{5.18}$$

This multilayer perceptron must be extended so as to satisfy the boundary conditions $y(0) = 1$ and $y(1) = 0$. In this regard, a set of possible particular and homogeneous solution terms is

$$\varphi_0(x) = (1 - x), \tag{5.19}$$
$$\varphi_1(x) = x(x - 1). \tag{5.20}$$

Equations (5.19) and (5.20) give

$$y(x;\underline{\alpha}) = (1-x) + x(x-1)y(x;\underline{\alpha}), \tag{5.21}$$

Also, the descent curve $y(x)$ must lie in the interval $(-\infty, 1)$, so the network outputs are bounded in the form

$$y(x;\underline{\alpha}) = \begin{cases} y(x;\underline{\alpha}), & y(x;\underline{\alpha}) < 1, \\ 1-\epsilon, & y(x;\underline{\alpha}) \geq 1, \end{cases} \tag{5.22}$$

being ϵ a small number. As we will see, this is a very delicate choice in this problem. Here we have use $\epsilon = 10^{-5}$.

On the other hand, the neural network biases and synaptic weights are initialized at random. Experience has shown that this direct method does not require a good initial guess for the solution in this problem, which is a positive result. Figure 5.9 shows the initial neural network response for this set of parameters chosen at random.

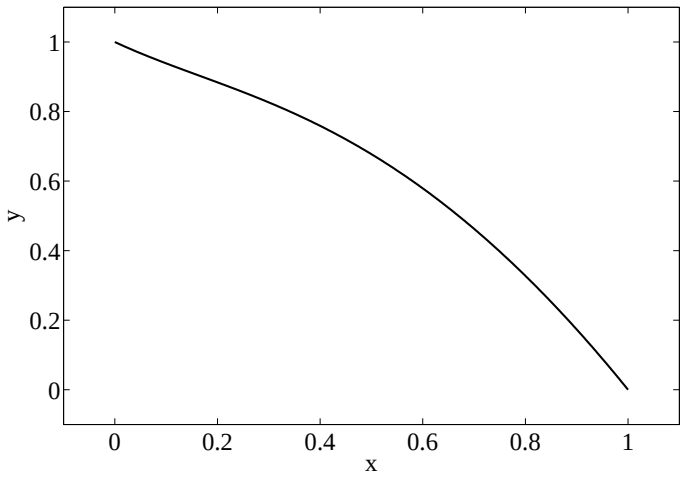


Figure 5.9: Initial guess for the brachistochrone problem.

5.2.3 Formulation of variational problem

Making use of Equation (5.15), the variational statement of the brachistochrone problem for the multilayer perceptron is to find a function $y^*(x;\underline{\alpha}^*) \in V$ for which the functional

$$F[y(x;\underline{\alpha})] = \frac{1}{\sqrt{2g}} \int_0^1 \sqrt{\frac{1+[y'(x;\underline{\alpha})]^2}{1-y(x;\underline{\alpha})}}\, dx, \tag{5.23}$$

defined on V, takes on a minimum value.

The derivative $y'(x;\underline{\alpha})$ can be evaluated analytically with the back-propagation algorithm for the Jacobian matrix, see Section 3.3.

On the other hand, and to evaluate the objective functional in Equation (5.23), the integration of a function is needed. Moreover, this objective functional has an integrable singularity at the lower limit. This singularity makes this problem very difficult to be solved numerically. One possibility here is to make

$$\int_{x_a}^{x_b} f(x)dx \approx \int_{x_a+\epsilon}^{x_b} f(x)dx, \tag{5.24}$$

being ϵ a small number. Here we use $\epsilon = 10^{-5}$. The integration is done with an ordinary differential equation approach and the Runge-Kutta method with 1001 integration points, see Sections B.1 and B.2.

Last, evaluation of the objective function gradient, $\nabla f(\underline{\alpha})$, is carried out by means of numerical differentiation. In particular, the central differences method is used with an ϵ value of 10^{-6} [15] .

5.2.4 Solution of reduced function optimization problem

A quasi-Newton method with BFGS train direction and Brent optimal train rate methods is the training algorithm chosen to solve the reduced function optimization problem. The tolerance in the Brent's method is set to 10^{-6}.

The objective function in this problem has shown to contain a single, the unique, global minimum. In contrast, other representations for the descent curve, such as piece-wise linear or polynomial, could lead to objective functions with many local minima. Indeed, convergence properties are related to the function space used.

The evaluation of the initial guess is 0.734351. The training algorithm is set to stop when it can not continue with the training process. This occurs after 237 epochs, when the Brent's method gives zero optimal train rate. Of course, a good initial guess would decrease the number of epochs. After training the evaluation falls to 0.595314. The gradient norm decreases from 0.197046 to $2.61258 \cdot 10^{-8}$. Figures 5.10 and 5.11 depict the evaluation value of the objective function and its gradient norm, versus the training epoch, respectively. Note that a logarithmic scale is used for the y-axis in the second plot. The evaluation history shows a strong reduction of the objective function during the first few epochs to become nearly constant until the stopping criterium is satisfied. On the other hand, the gradient norm decreases in several orders of magnitude during the training process to reach a very small value. These two situations show that the quasi-Newton method has certainly converged.

Table 5.2 shows the training results for this problem. Here N is the number of training epochs, M the number of objective function evaluations, CPU the CPU time in seconds in a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameter vector norm, $t(\underline{\alpha}^*)$ the final evaluation value and $\nabla f(\underline{\alpha}^*)$ the final gradient norm. The descent time provided by the neural network yields a percentage error of around -3.0764% when compared to the analytical solution, being the singularity in the integral (5.23) the main cause of error. Anyway, results here are good, since the descent time provided by the network is quite similar to the descent time for the brachistochrone.

The explicit form of the function represented by the trained neural network is

$$\begin{aligned} y^*(x;\underline{\alpha}^*) &= (1-x) + x(x-1)[38.7381 \\ &- 59.6973 \tanh(2.19639 + 1.57101x) \\ &+ 21.9216 \tanh(48.4807 + 120.955x) \\ &+ 0.392256 \tanh(0.990452 - 14.656x)]. \end{aligned} \tag{5.25}$$

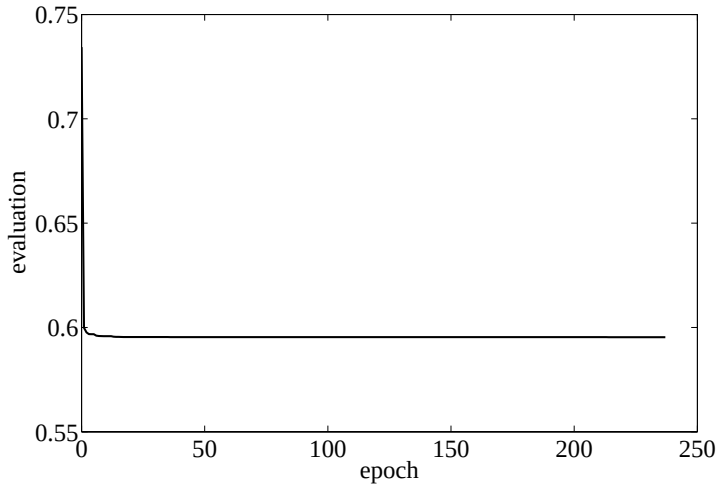


Figure 5.10: Evaluation history for the brachistochrone problem.

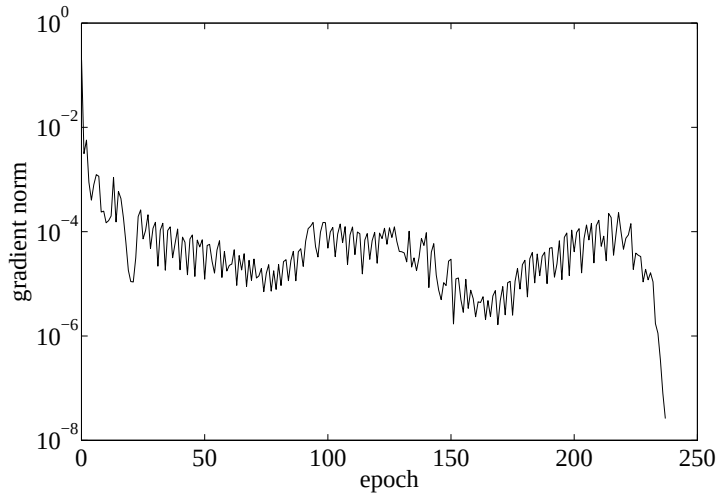


Figure 5.11: Gradient norm history for the brachistochrone problem.

for $x \in [0, 1]$. Figure 5.12 illustrates the neural network solution. This shape also agrees to that of the cycloid, which is the analytical solution to the brachistochrone problem.

5.3 The catenary problem

The catenary problem for the multilayer perceptron is a variational problem with one input and one output variables, two boundary conditions, and an objective functional with one constraint and requiring the integration of two functions. This problem is implemented in the C++ class

N	=	237
M	=	13526
CPU	=	$1166s$
$\|\underline{\alpha}^*\|$	=	47.6955
$t(\underline{\alpha}^*)$	=	0.595314
$\|\nabla t(\underline{\alpha}^*)\|$	=	$2.61258 \cdot 10^{-8}$

Table 5.2: Training results for the brachistochrone problem.

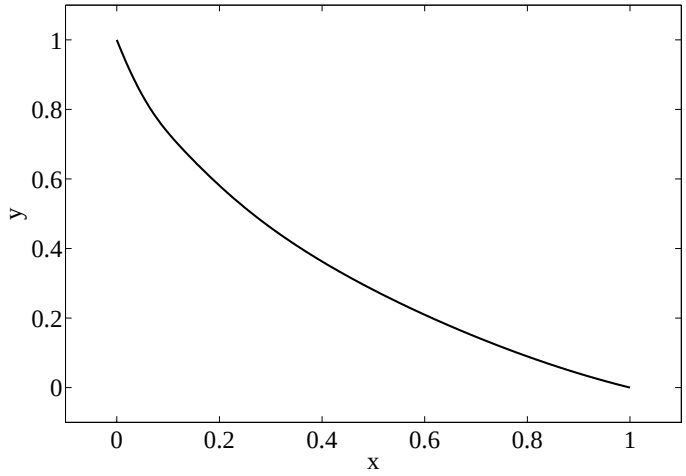


Figure 5.12: Neural network results for the brachistochrone problem.

library Flood [61].

5.3.1 Problem statement

The catenary problem is posed as follows:

To find the curve assumed by a loose string of length l hung freely from two fixed points $A = (x_a, f_a)$ and $B = (x_b, f_b)$.

Figure 5.13 graphically declares the catenary problem.
The length of a chain $y(x)$ is given by

$$L[y(x)] = \int_{x_a}^{x_b} \sqrt{1 + [y'(x)]^2} dx. \tag{5.26}$$

This chain is constrained to have length l, which can be written

$$E_L[y(x)] = \int_{x_a}^{x_b} \sqrt{1 + [y'(x)]^2} dx - l$$

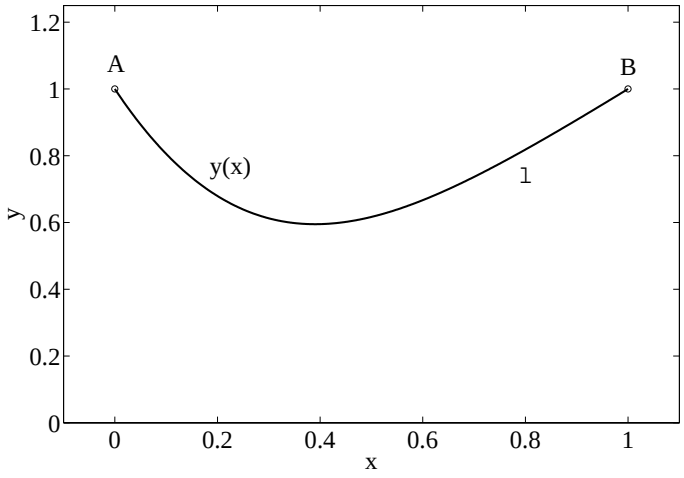


Figure 5.13: The catenary problem statement.

$$= \quad 0. \tag{5.27}$$

On the other hand, the shape to be found is that which minimizes the potential energy. For a chain $y(x)$ with uniformly distributed mass this is given by

$$V[y(x)] = \int_{x_a}^{x_b} y(x)\sqrt{1 + [y'(x)]^2}dx. \tag{5.28}$$

The analytical solution to the catenary problem is an hyperbolic cosine. For the particular case when $l = 1.5$, $A = (0, 1)$ and $B = (1, 1)$, it is written

$$y^*(x) = 0.1891 + 0.3082 \cosh\left(\frac{x - 0.5}{0.3082}\right). \tag{5.29}$$

The potential energy of this catenary is $V[y^*(x)] = 1.0460$.

5.3.2 Selection of function space

A multilayer perceptron with one input, three sigmoid neurons in the hidden layer and a linear output neuron is used to span a function space for the catenary problem. Here the input represents the independent variable x, while the output represents the dependent variable y, see Figure 5.14.

The family of functions created by this neural network is denoted V and it is of dimension $s = 10$. The functions belonging to V are of the type

$$\begin{array}{rccc} y: & \mathbb{R} & \rightarrow & \mathbb{R} \\ & x & \mapsto & y(x; \underline{\alpha}), \end{array}$$

where

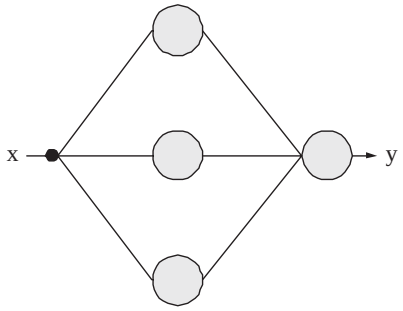


Figure 5.14: Network architecture for the catenary problem.

$$y(x;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} x\right). \tag{5.30}$$

However, this is not a suitable starting point for the catenary problem, since the elements of V do not hold the boundary conditions $y(0;\underline{\alpha}) = 1$ and $y(1;\underline{\alpha}) = 1$. For that reason, the multilayer perceptron of Equation (5.30) must be extended with a particular and an homogeneous solution terms such as

$$\varphi_0(x) = 1, \tag{5.31}$$
$$\varphi_1(x) = x(x-1), \tag{5.32}$$

to give

$$y(x;\underline{\alpha}) = 1 + x(x-1)y(x;\underline{\alpha}). \tag{5.33}$$

Now all the functions of the family defined by Equation (5.33) satisfy $y(0;\underline{\alpha}) = 1$ and $y(1;\underline{\alpha}) = 1$.

The neural network must be initialized by assigning some values to the parameters. Here we adopt a random initial guess, see Figure 5.15.

5.3.3 Formulation of variational problem

From Equations (5.27) and (5.28), the constrained problem can be stated as to find a function $y^*(x;\underline{\alpha}^*) \in V$ such that

$$\begin{aligned} E_L[y^*(x;\underline{\alpha}^*)] &= \int_{x_a}^{x_b} \sqrt{1 + [y'(x;\underline{\alpha}^*)]^2}\,dx - l \\ &= 0 \end{aligned} \tag{5.34}$$

and for which the functional

$$V[y(x;\underline{\alpha})] = \int_{x_a}^{x_b} y(x;\underline{\alpha})\sqrt{1 + [y'(x;\underline{\alpha})]^2}\,dx, \tag{5.35}$$

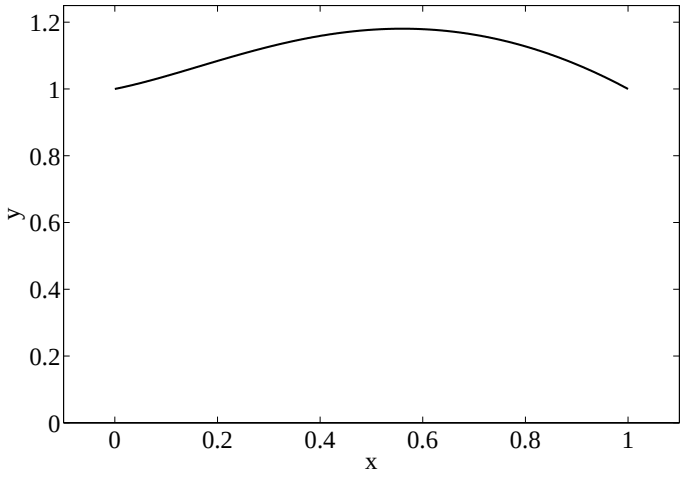


Figure 5.15: Initial guess for the catenary problem.

defined on V, takes on a minimum value.

The use of a penalty term for the constraint in the length can approximate this problem to an unconstrained one, which is now formulated as to find a function $y^*(x;\underline{\alpha}^*) \in V$ for which the functional

$$F[y(x;\underline{\alpha})] = \rho_L(E_L[y(x;\underline{\alpha}^*)])^2 + \rho_V V[y(x;\underline{\alpha}^*)], \tag{5.36}$$

defined on V, takes on a minimum value.

The numbers ρ_L and ρ_V can be named length error term weight and potential energy term weight, respectively. These are design variables, and their values should be chosen so that the constraint term dominates over the objective term. Here we have set $\rho_L = 100$ and $\rho_V = 1$.

The value of $y'(x;\underline{\alpha})$ in Equations (5.34) and (5.35) is obtained by means of the back-propagation algorithm for the Jacobian matrix. This is derived in Section 3.3.

On the other hand, evaluation of (5.36) requires the integration of two functions. Here we use an ordinary differential equation approach to approximate the integrals with the Runge-Kutta-Fehlberg method, see Sections B.1 and B.2. The tolerance is set to 10^{-15}.

As it is not possible to derive a back-propagation algorithm for the gradient $\nabla f(\underline{\alpha})$ in this problem, that item is obtained numerically with central differences and using $\epsilon = 10^{-6}$

5.3.4 Solution of reduced function optimization problem

A quasi-Newton method with BFGS inverse Hessian approximation and Brent's line minimization methods is used here for training. No stopping criterion is chosen. Instead training will finish when the algorithm can't keep on doing. At that point the Brent's method gives zero train rate. Figures 5.16 and 5.17 depict the training histories for the evaluation and the gradient norm. They both supply good signs about the this process, since the evaluation and the gradient norm decrease significantly.

Some results by the quasi-Newton method are listed in Table 5.3. This contains number of epochs N, number of evaluations M, CPU time in a laptop AMD 3000 CPU, final parameters

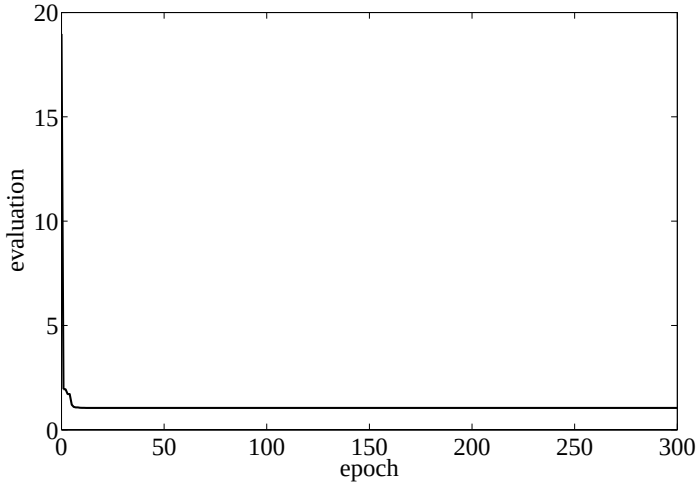


Figure 5.16: Evaluation history for the catenary problem.

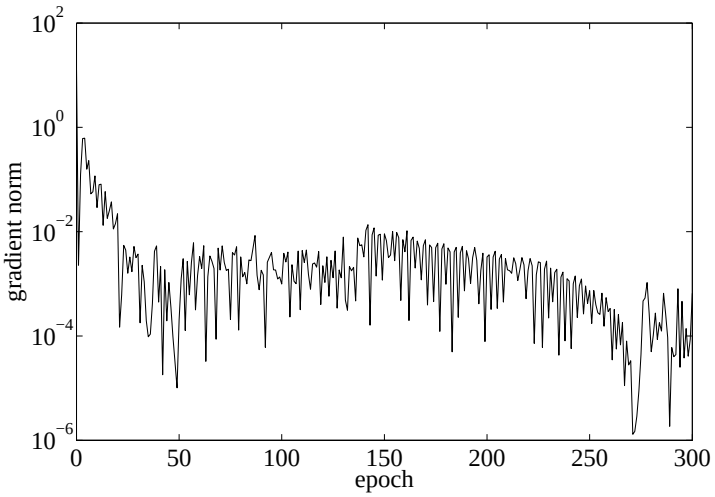


Figure 5.17: Gradient norm history for the catenary problem.

norm $\|\underline{\alpha}^*\|$, final evaluation $f(\underline{\alpha}^*)$, final length error $e_l(\underline{\alpha}^*)$, final potential energy $v(\underline{\alpha}^*)$ and final gradient norm $\|\nabla f(\underline{\alpha}^*)\|$.

The multilayer perceptron for the catenary problem can be written explicitly as

$$\begin{aligned} y^*(x;\underline{\alpha}^*) &= 1 + x(x-1)[0.417922 \\ &- 5.13772 \tanh\left(0.390226 + 0.507345x\right) \\ &+ 2.5535 \tanh\left(-1.24642 + 1.08916x\right) \end{aligned}$$

N	=	300
M	=	17328
CPU	=	$871s$
$\|\underline{\alpha}^*\|$	=	3.05481
$f(\underline{\alpha}^*)$	=	1.04588
$e_l(\underline{\alpha}^*)$	=	$9.53529 \cdot 10^{-4}$
$v(\underline{\alpha}^*)$	=	1.04579
$\|\nabla f(\underline{\alpha}^*)\|$	=	$6.60019 \cdot 10^{-4}$

Table 5.3: Training results for the catenary problem.

$$+ \quad 6.05167 \tanh\left(4.20948 + 1.65389x\right)], \tag{5.37}$$

for $x \in [0, 1]$.

To conclude, a plot of Equation (5.37) is shown in Figure 5.18. Here we can see a good match between the neural network results and the analytical catenary.

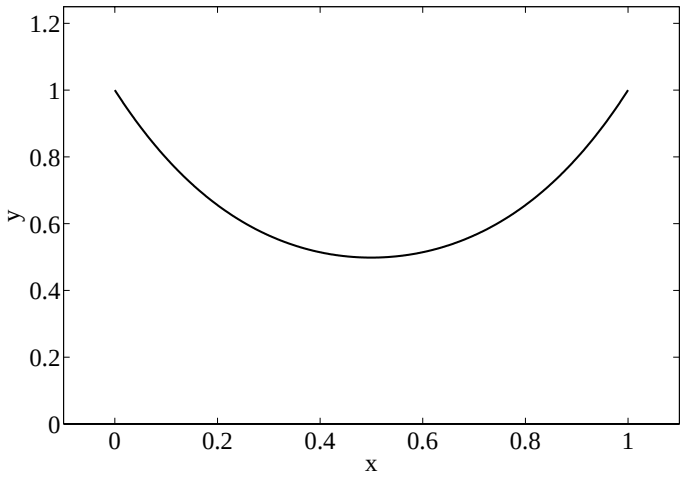


Figure 5.18: Neural network results for the catenary problem.

5.4 The isoperimetric problem

The isoperimetric problem for the multilayer perceptron is a variational problem with one input and two output variables, four boundary conditions and one constraint in the objective functional, which is evaluated by integrating two functions. All the code in the C++ programming language needed to solve this problem can be found in Flood [61]. Another variation of the isoperimetric problem for the multilayer perceptron has been published in [66].

5.4.1 Problem statement

The isoperimetric problem is another classical problem in the calculus of variations [99]. The statement of this problem is:

Of all simple closed curves in the plane of a given length l, which encloses the maximum area?.

Figure 5.19 is a graphical statement of the isoperimetric problem. Note that we can not use here a function $y(x)$ to specify the closed curve, since closed curves will necessarily make the function multi-valued. Instead, we use the parametric equations $(x, y)(t)$, for $t \in [0, 1]$, and such that $x(0) = 0$, $x(1) = 0$, $y(0) = 0$ and $y(1) = 0$.

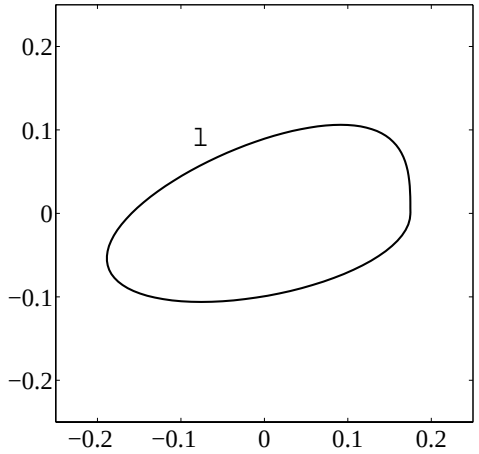


Figure 5.19: The isoperimetric problem statement.

For a plane curve specified in parametric equations as $(x, y)(t)$, the arc length is given by [99]

$$L[(x,y)(t)] = \int_0^1 \sqrt{[x'(t)]^2 + [y'(t)]^2}dt. \tag{5.38}$$

From Equation (5.38), the constraint on the length can be expressed as an error functional,

$$\begin{aligned} E_L[(x,y)(t)] &= \int_0^1 \sqrt{[x'(t)]^2 + [y'(t)]^2}dt - l \\ &= 0. \end{aligned} \tag{5.39}$$

On the other hand, Green's theorem gives the signed area as [99]

$$A[(x,y)(t)] = \frac{1}{2}\int_0^1 \left(x(t)y'(t) - x'(t)y(t)\right) dt. \tag{5.40}$$

The analytical solution to the isoperimetric problem is a circle [99]. Taking, for instance, the perimeter goal to be $l = 1$, the optimal function is given by

$$x(t) = -\frac{1}{2\pi} + \frac{1}{2\pi}\cos(2\pi t), \tag{5.41}$$

$$y(t) = \frac{1}{2\pi}\sin(2\pi t), \tag{5.42}$$

for $t \in [0, 1]$. That circle yields a maximum area $A[(x, y)^*(t)] = 1/(4\pi)$. The analytical solution to this problem can be found in Section 2.3.

5.4.2 Selection of function space

A neural network consisting of a hidden layer of sigmoid perceptrons and an output layer of linear perceptrons is used here to represent the closed curve. As this shape is written in parametric equations as $(x, y) = (x, y)(t)$, the network must have one input, t, and two output neurons, x and y. The number of neurons in the hidden layer is up to the designer. Anyhow, this is not a very important choice in this problem, since it can be regarded well-conditioned. Therefore, a neural network with sufficient complexity will be enough. Here we use three neurons in the hidden layer. Figure 5.20 depicts this network architecture.

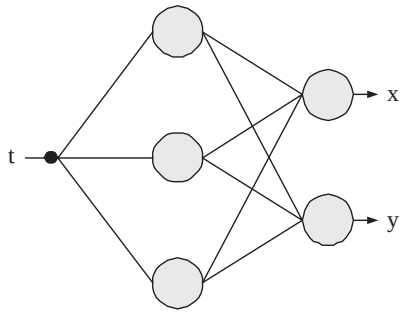


Figure 5.20: Network architecture for the isoperimetric problem.

The neural network in Figure 5.20 spans a function space V of dimension $s = 14$. The elements of this family of functions are of the form

$$\begin{aligned} (x, y): \quad \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto (x, y)(t; \underline{\alpha}), \end{aligned}$$

where

$$x(t; \underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right), \tag{5.43}$$

$$y(t; \underline{\alpha}) = b_2^{(2)} + \sum_{j=1}^{3} w_{2j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right). \tag{5.44}$$

The function space for this problem is required to satisfy the boundary conditions $x(0) = 0$, $x(1) = 0$, $y(0) = 0$ and $y(1) = 0$. A set of homogeneous and particular solution terms here could be

$$\varphi_{x0}(t) = 0, \tag{5.45}$$
$$\varphi_{x1}(t) = t(t-1), \tag{5.46}$$
$$\varphi_{y0}(t) = 0, \tag{5.47}$$
$$\varphi_{y1}(t) = t(t-1). \tag{5.48}$$

Elements of V thus become

$$x(t;\underline{\alpha}) = t(t-1)x(t;\underline{\alpha}), \tag{5.49}$$
$$y(t;\underline{\alpha}) = t(t-1)y(t;\underline{\alpha}). \tag{5.50}$$

Please note that although (5.49) is written as two different equations they represent just one function, where many of the parameters are shared by both output variables. This is a potential advantage over other direct methods which might use as many different functions as output variables. Indeed, if the outputs are related in some way, shared parameters lead in a reduction in the dimension of the function space.

Finally, the neural network is initialized with parameters chosen at random. Figure (5.21) shows the initial guess for the isoperimetric problem considered here. As the initial solution is a random shape, it does not necessarily satisfy the constraint on the length.

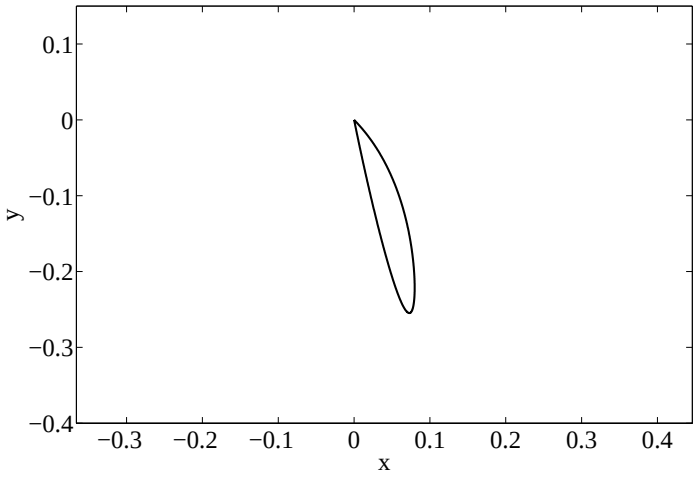


Figure 5.21: Initial guess for the isoperimetric problem.

5.4.3 Formulation of variational problem

Considering Equations (5.39) and (5.40) the variational statement of the isoperimetric problem for the multilayer perceptron is to find a function $(x, y)^*(t;\underline{\alpha}^*) \in V$, for $t \in [0, 1]$, such that

$$\begin{aligned} E_L[(x,y)^*(t;\underline{\alpha}^*)] &= \int_0^1 \sqrt{[x'(t)]^2 + [y'(t)]^2}dt - l \\ &= 0, \end{aligned} \tag{5.51}$$

and for which the functional

$$A[(x,y)(t;\underline{\alpha})] = \frac{1}{2}\int_0^1 (x(t;\underline{\alpha})y'(t;\underline{\alpha}) - x'(t;\underline{\alpha})y(t;\underline{\alpha}))\,dt, \tag{5.52}$$

defined on V, takes on a maximum value.

This constrained problem can be reformulated as an unconstrained problem by adding a penalty term for the constraint E_L to the original objective functional A. The reduced problem is now stated as to find a function $(x,y)^*(t;\underline{\alpha}^*) \in V$, for $t \in [0,1]$, and for which the objective functional

$$F[(x,y)(t;\underline{\alpha})] = \rho_L(E_L[(x,y)(t;\underline{\alpha})])^2 - \rho_A A[(x,y)(t;\underline{\alpha})], \tag{5.53}$$

defined on V, takes on a minimum value.

Here ρ_L is the length error term weight and ρ_A is the area term weight. Both parameters are design variables in the problem, and should be chosen so that the initial value of the objective functional is of order unity and so that the values of the penalty term is much smaller than the value of the objective term. For this problem we have chosen $\rho_L = 100$ and $\rho_A = 1$.

The quantities $x'(t;\underline{\alpha})$ and $y'(t;\underline{\alpha})$ are computed with the back-propagation algorithm for the derivatives of the output variables with respect to the input variables, i.e., the Jacobian matrix.

On the other hand, evaluation of (5.53) requires the integration of two functions, which is performed here numerically. An ordinary differential equation approach using an adaptive stepsize algorithm gives the greatest accuracy. In particular we use the Runge-Kutta-Fehlberg method with a tolerance of 10^{-15}.

Finally, the gradient vector of the objective function, $\nabla f(\underline{\alpha}^*)$, is computed numerically applying central differences with $\epsilon = 10^{-6}$.

5.4.4 Solution of reduced function optimization problem

A quasi-Newton method with BFGS train direction and Brent optimal train rate is adopted here for solving the reduced function optimization problem. This particular objective function has demonstrated not to contain local minima. Certainly, the function space spanned by the neural network has good convergence properties. On the contrary, other representations for the closed curve, such as piece-wise linear or polynomial, could lead to objective functions with many local minima.

In this example, no more reduction in the objective function will cause to finish the training process. It might happen that the Brent's method gives zero train rate for some BFGS train direction. In such occasions we can restart the training process. On the other hand, constrained problems need in general more training effort that unconstrained problems. Also, a good initial guess can decrease very much that computing activity. Figure 5.22 shows a evaluation history which decreases very rapidly to become almost constant, while Figure 5.23 depicts a gradient norm history decreasing in several orders of magnitude and with a very small final value. Therefore, we can say that the quasi-Newton method has found a minimum.

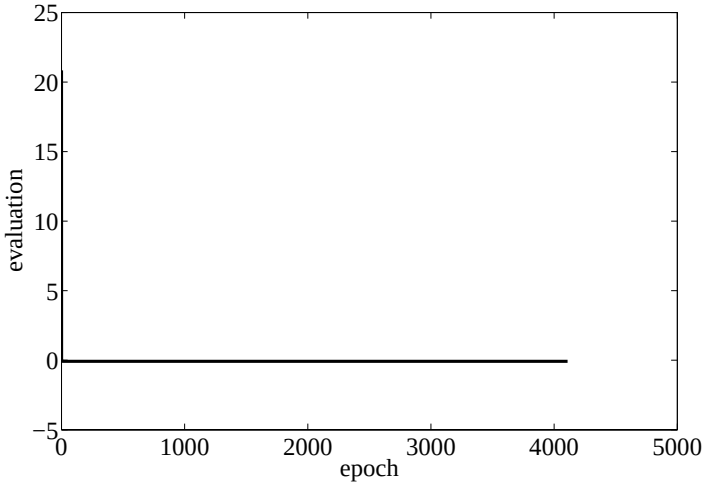


Figure 5.22: Evaluation history for the isoperimetric problem.

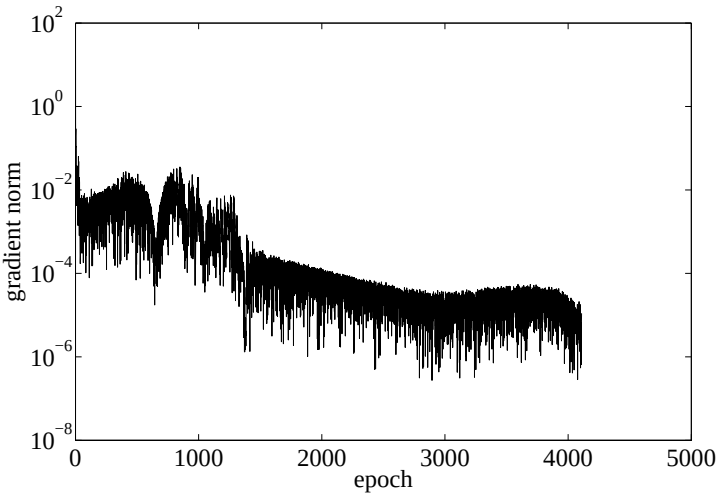


Figure 5.23: Gradient norm history for the isoperimetric problem.

Table 5.4 shows the training results for this problem. Here N is the number of training epochs, M the number of evaluations, CPU the training time for a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameters norm $f(\underline{\alpha}^*)$ the final objective function value, $e_l(\underline{\alpha}^*)$ the final length error, $a(\underline{\alpha}^*)$ the final closed curve area and $\|\nabla f(\underline{\alpha}^*)\|$ the final gradient norm. The perimeter error is less than 10^{-3} and the error in the closed curve by the neural network is 0.1256%. Results here are also good, since the perimeter error is very small and the neural network area is very similar to that of the circle.

The explicit expression of the function addressed by this neural network is written

N	$=$	4109
M	$=$	138507
CPU	$=$	$24044s$
$\|\underline{\alpha}^*\|$	$=$	11.0358
$f(\underline{\alpha}^*)$	$=$	-0.0796397
$e_l(\underline{\alpha}^*)$	$=$	$7.96674 \cdot 10^{-4}$
$a(\underline{\alpha}^*)$	$=$	0.0797032
$\|\nabla f(\underline{\alpha}^*)\|$	$=$	$1.10828 \cdot 10^{-5}$

Table 5.4: Training results for the isoperimetric problem.

$$
\begin{aligned}
x^*(t;\underline{\alpha}^*) &= t(t-1)[10.7483 \\
&+ 4.68016\tanh\left(-4.05158 + 2.43348t\right) \\
&- 12.4708\tanh\left(-0.196899 + 0.680774t\right) \\
&+ 14.4052\tanh\left(-0.789403 + 0.839184t\right)], \qquad (5.54)\\
y^*(t;\underline{\alpha}^*) &= t(t-1)[-23.7525 \\
&- 24.3861\tanh\left(-4.05158 + 2.43348t\right) \\
&+ 4.62781\tanh\left(-0.196899 + 0.680774t\right) \\
&+ 0.374342\tanh\left(-0.789403 + 0.839184t\right)], \qquad (5.55)
\end{aligned}
$$

for $t \in [0, 1]$ and where all the parameter values have been rounded to the third decimal point.

To end, Figure 5.24 illustrates the neural network solution to the isoperimetric problem. We can see that the shape provided by the numerical method is that of the circle.

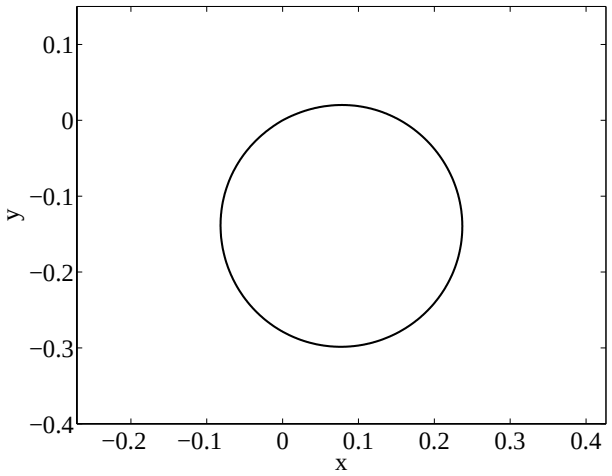


Figure 5.24: Neural network results for the isoperimetric problem.

This page is intentionally left blank.

Chapter 6

Optimal control problems

This section gives a brief introduction to optimal control theory, in which the most important aspects of mathematical formulation are reviewed. The problem of optimal control is intimately related to the calculus of variations, and therefore it can be formulated as a novel learning task for the multilayer perceptron.

In order to validate this numerical method, an optimal control problem with analytical solution is solved by means of a multilayer perceptron. Two practical applications of chemical and aeronautical engineering interest are also approached with encouraging results.

6.1 Problem formulation

Optimal control -which is playing an increasingly important role in the design of modern systems- has as its objective the optimization, in some defined sense, of physical processes. More specifically, the objective of optimal control is to determine the control signals that will cause a process to satisfy the physical constraints and at the same time minimize or maximize some performance criterion [53].

The formulation of an optimal control problem requires:

(i) A mathematical model of the system.

(ii) A statement of the physical constraints.

(iii) A specification of the performance criterion.

Mathematical model

The model of a process is a mathematical description that adequately predicts the response of the physical system to all anticipated inputs.

Let denote $\mathbf{y}(\mathbf{x})$ the control variables to the system and $\mathbf{u}(\mathbf{x})$ the state variables of the system. Then the mathematical model (or state equation) can be written as

$$\mathcal{L}(\mathbf{y}(\mathbf{x}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}, \tag{6.1}$$

where $\mathcal{L}$ is some algebraic or differential operator and $\mathbf{f}$ is some forcing term.

Physical constraints

An optimal control problem might be specified by a set of constraints on the controls and the states of the system.

Two important types of control constraints are boundary conditions and lower and upper bounds. If some outputs are specified for given inputs, then the problem is said to include boundary conditions. On the other hand, if some control variables are restricted to fall in some interval, then the problem is said to have lower and upper bounds.

State constraints are conditions that the physical system must satisfy. This type of constraints vary according to the problem at hand.

In this way, a control which satisfies all the control and state constraints is called an admissible control [53].

Definition 10 (Admissible control)**.** A control $\mathbf{y}(\mathbf{x})$ which satisfies all the constraints is called an admissible control. The set of admissible controls is denoted Y, and the notation $\mathbf{y}(\mathbf{x}) \in Y$ means that $\mathbf{y}(\mathbf{x})$ is admissible.

Similarly, a state which satisfies the state constraints is called an admissible state [53].

Definition 11 (Admissible state)**.** A state $\mathbf{u}(\mathbf{x})$ which satisfies the state variable constraints is called an admissible state. The set of admissible states will be denoted by $\mathbf{U}$, and $\mathbf{u}(\mathbf{x}) \in \mathbf{U}$ means that the state $\mathbf{u}(\mathbf{x})$ is admissible.

Performance criterion

In order to evaluate the performance of a system quantitatively, a criterion must be chosen. The performance criterion is a scalar functional of the control variables of the form

$$\begin{array}{rccc} F: & Y & \to & \mathbb{R} \\ & \mathbf{y}(\mathbf{x}) & \mapsto & F[\mathbf{y}(\mathbf{x})]. \end{array}$$

In certain cases the problem statement might clearly indicate which performance criterion is to be selected, whereas in other cases that selection is a subjective matter [53].

An optimal control is defined as one that minimizes or maximizes the performance criterion, and the corresponding state is called an optimal state. More specifically, the optimal control problem can be formulated as

Problem 7 (Optimal control problem)**.** Let Y and U be the function spaces of admissible controls and states, respectively. Find an admissible control $\mathbf{y}^*(\mathbf{x}) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to be in an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$ and for which the performance criterion

$$F[\mathbf{y}(\mathbf{x})]$$

takes on a minimum or maximum value. The function $\mathbf{y}^*(\mathbf{x})$ is called an optimal control and the function $\mathbf{u}^*(\mathbf{x})$ an optimal state.

In this way, the problem of optimal control is formulated as a variational problem [53].

Solution approach

In general, optimal control problems lead to a variational problem that cannot be solved analytically to obtain the optimal control signal. In order to achieve this goal, two types of numerical methods are found in the literature, namely, direct and indirect [14]. From them, direct methods are the most widely used.

As it has been explained in this PhD Thesis, a variational formulation for neural networks provides a direct method for the solution of variational problems. Therefore optimal control problems can be approached with this numerical technique. Following the nomenclature introduced in this section, the optimal control problem for the multilayer perceptron translates as follows:

Problem 8 (Optimal control problem for the multilayer perceptron)**.** Let Y and U be the function spaces of admissible controls and states, respectively. Let also V be the space consisting of all controls $\mathbf{y}(\mathbf{x};\underline{\alpha})$ that a given multilayer perceptron can define, with dimension s. Find an admissible control $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x};\underline{\alpha}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to be in an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$, and for which the performance criterion

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})],$$

defined on V, takes on a minimum or a maximum value.

6.2 Validation examples

In this section a variational formulation for the multilayer perceptron is applied for validation to an optimal control problem. The selected example can be solved analytically, which enables a fair comparison with the neural network results.

6.2.1 The car problem

The car problem for the multilayer perceptron is an optimal control problem with two controls and two state variables. It is defined by an objective functional with two constraints and requiring the integration of a system of ordinary differential equations. This problem is included with the Flood library [61], and it has been published in [63]. Another variant of the car problem not considered here can also be found in [68].

Problem statement

Consider a car which is to be driven along the x-axis from some position x_i at velocity v_i to some desired position x_f at desired velocity v_f in a minimum time t_f, see Figure 6.1.

To simplify the problem, let us approximate the car by a unit point mass that can be accelerated by using the throttle or decelerated by using the brake. Selecting position and velocity as state variables the mathematical model of this system becomes a Cauchy problem of two ordinary differential equations with their corresponding initial conditions,

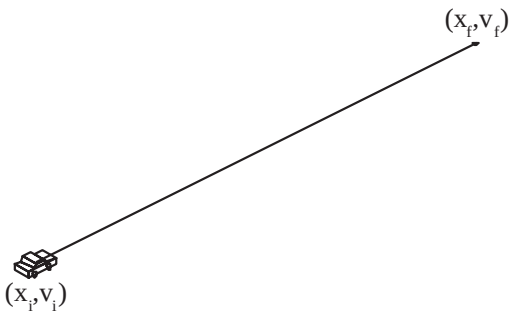


Figure 6.1: The car problem statement.

$$\begin{aligned} \dot{x}(t) &= v(t), && (6.2)\\ \dot{v}(t) &= a(t) + d(t), && (6.3)\\ x(0) &= x_i, && (6.4)\\ v(0) &= v_i, && (6.5) \end{aligned}$$

for $t \in [0, t_f]$ and where the controls $a(t)$ and $d(t)$ are the throttle acceleration and the braking deceleration, respectively.

The acceleration is bounded by the capability of the engine, and the deceleration is limited by the braking system parameters. If the maximum acceleration is $\sup(a) > 0$, and the maximum deceleration is $\sup(d) > 0$, such bounds on the control variables can be written

$$\begin{aligned} 0 \leq \quad a(t) \quad &\leq \sup(a), && (6.6)\\ -\sup(d) \leq \quad d(t) \quad &\leq 0. && (6.7) \end{aligned}$$

As the objective is to make the car reach the final point as quickly as possible, the objective functional for this problem is given by

$$F[(a,d)(t)] = t_f. \tag{6.8}$$

On the other hand, the car is to be driven to a desired position x_f and a desired velocity v_f, therefore $x(t_f) = x_f$ and $v(t_f) = v_f$. Such constraints on the state variables can be expressed as error functionals,

$$\begin{aligned} E_x[(a,d)(t)] &= x(t_f) - x_f \\ &= 0, && (6.9) \end{aligned}$$

$$\begin{aligned} E_v[(a,d)(t)] &= v(t_f) - v_f \\ &= 0. && (6.10) \end{aligned}$$

where E_x and E_v are called the final position and velocity errors, respectively.

If we set the initial position, initial velocity, final position, final velocity, maximum acceleration and maximum deceleration to be $x_i = 0$, $v_i = 0$, $x_f = 1$, $v_f = 0$, $\sup(a) = 1$ and $\sup(d) = 1$, respectively. This problem has an analytical solution for the optimal control given by [53]

$$a^*(t) = \begin{cases} 1, & 0 \leq t < 1, \\ 0, & 1 \leq t \leq 2, \end{cases} \tag{6.11}$$

$$d^*(t) = \begin{cases} 0, & 0 \leq t < 1, \\ -1, & 1 \leq t \leq 2, \end{cases} \tag{6.12}$$

which provides a minimum final time $t_f^* = 2$.

The statement and the solution itself of this car problem points out a number of significant issues. First, some variational problems might require a function space with independent parameters associated to it. Indeed, the final time is not part of the control, but it represents the interval when it is defined. Finally, this kind of applications demand spaces of functions with very good approximation properties, since they are likely to have very non-linear solutions. Here the optimal control even exhibits discontinuities.

Selection of function space

Here a multilayer perceptron with a sigmoid hidden layer and a linear output layer is chosen to represent the control $(a, b)(t)$. The neural network must have one input, t, and two output neurons, a and d. Although the size of the hidden layer is a design variable, that number is not a critical issue in optimal control. Indeed, this class of problems are not regarded as being ill-posed, and a sufficient complexity for the function space selected is generally enough. In this problem we use three hidden neurons. Figure 6.2 is a graphical representation of this network architecture.

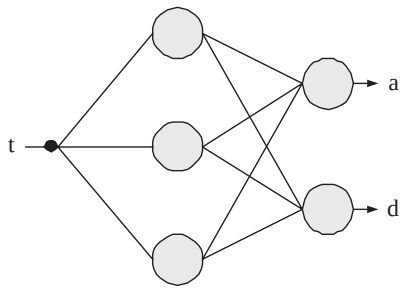


Figure 6.2: Network architecture for the car problem.

Also this neural network needs an associated independent parameter representing the final time t_f.

Such a multilayer perceptron spans a family V of parameterized functions $(a, b)(t; \underline{\alpha})$ of dimension $s = 14 + 1$, being 14 the number of biases and synaptic weights and 1 the number of independent parameters. Elements V are of the form

$$\begin{aligned} (a, d): \quad \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto (a, d)(t; \underline{\alpha}, t_f), \end{aligned}$$

where

$$a(t;\underline{\alpha},t_f) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right), \tag{6.13}$$

$$d(t;\underline{\alpha},t_f) = b_2^{(2)} + \sum_{j=1}^{3} w_{2j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right). \tag{6.14}$$

Equation (6.13) represents just one function, in many of the parameters are shared.

The control variable is constrained to lie in the interval $[0, 1]$. To deal with such constraints we bound the network outputs in the form

$$a(t;\underline{\alpha},t_f) = \begin{cases} 0, & a(t;\underline{\alpha},t_f) < 0. \\ a(t;\underline{\alpha},t_f), & 0 \leq a(t;\underline{\alpha},t_f) \leq 1. \\ 1, & a(t;\underline{\alpha},t_f) > 1. \end{cases} \tag{6.15}$$

$$d(t;\underline{\alpha},t_f) = \begin{cases} -1, & d(t;\underline{\alpha},t_f) < -1. \\ d(t;\underline{\alpha},t_f), & -1 \leq d(t;\underline{\alpha},t_f) \leq 0. \\ 0, & d(t;\underline{\alpha},t_f) > 0. \end{cases} \tag{6.16}$$

All the parameters in the neural network are initialized at random.

Formulation of variational problem

From Equations (6.9), (6.10) and (6.8), the car problem formulated in this Section can be stated so as to find a control $(a,d)^*(t;\underline{\alpha}^*,t_f^*) \in V$ such that

$$E_x[(a,d)^*(t;\underline{\alpha}^*,t_f^*)] = 0, \tag{6.17}$$

$$E_v[(a,d)^*(t;\underline{\alpha}^*,t_f^*)] = 0, \tag{6.18}$$

and for which the functional

$$T[(a,d)(t;\underline{\alpha},t_f)],$$

defined on V, takes on a minimum value.

This constrained problem can be converted to an unconstrained one by the use of penalty terms. The statement of this unconstrained problem is now to find a control $(a,d)^*(t;\underline{\alpha}^*,t_f^*)$ for which the objective functional

$$F[(a,d)(t;\underline{\alpha},t_f)] = \rho_X E_x^2 + \rho_V E_v^2 + \rho_T T, \tag{6.19}$$

defined on V, takes on a minimum value.

The values $\rho_X = 10^{-3}$, $\rho_V = 10^{-3}$ and $\rho_T = 1$ are called the final time, error position an error velocity term weights, respectively.

Please note that evaluation of the objective functional (6.19) requires a numerical method for integration of ordinary differential equations. Here we choose the Runge-Kutta-Fehlberg method with tolerance 10^{-12} [93].

The objective function gradient vector, $\nabla f(\underline{\alpha})$, must be evaluated with numerical differentiation. In particular, we use the symmetrical central differences method [15] with an epsilon value of 10^{-6}.

Solution of reduced function optimization problem

Here we use a quasi-Newton method with BFGS train direction and Brent optimal train rate methods [83]. The tolerance in the Brent's method is set to 10^{-6}. While other direct methods might suffer from local optima with that algorithm in this problem, the neural networks method proposed here has demonstrated fully convergence to the global optimum.

In this example, training is stopped when the Brent's method gives zero rate for some gradient descent direction. The evaluation of the initial guess was 0.827; after 112 epochs of training this value falls to $1.999 \cdot 10^{-3}$.

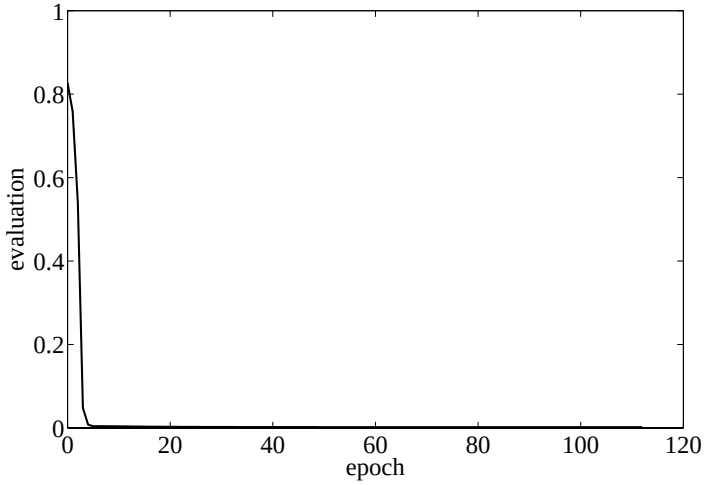


Figure 6.3: Evaluation history for the car problem.

Table 6.1 shows the training results for this problem. Here N denotes the number of epochs, M the number of evaluations, F the final objective functional value, ∇f the final objective function gradient norm, E_x the final position error, E_v the final velocity error and t_f^* the optimum time. As we can see, the final errors in the position and the velocity of the car are very small, and the final time provided by the neural network matches the analytical final time provided by the optimal function in Equation (6.11). More specifically, the errors made in the constraints are around $5 \cdot 10^{-3}$ and the error made in the final time is around 0.2%.

The analytical form of the optimal control addressed by the neural network is as follows

$$\begin{aligned} a^*(\underline{\alpha}^*, t_f^*) &= -1.31175 \\ &+ 6.85555 \tanh(-1.1448 + 1.48771t) \\ &- 0.387495 \tanh(2.52653 - 1.5223t) \\ &+ 16.1508 \tanh(12.2927 - 12.3053t), \end{aligned} \tag{6.20}$$

$$\begin{aligned} d^*(\underline{\alpha}^*, t_f^*) &= 1.82681 \\ &- 4.91867 \tanh(-1.1448 + 1.48771t) \\ &- 0.839186 \tanh(2.52653 - 1.5223t) \\ &+ 6.76623 \tanh(12.2927 - 12.3053t), \end{aligned} \tag{6.21}$$

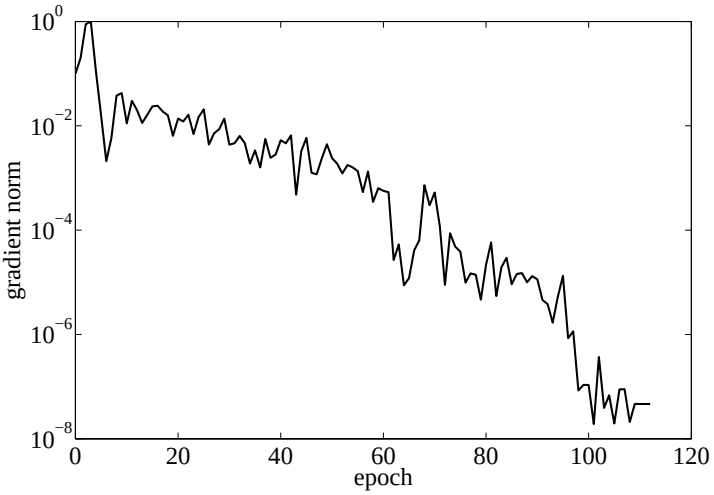


Figure 6.4: Gradient norm history for the car problem.

N	=	113
M	=	7565
CPU	=	$324s$
$\|\underline{\alpha}^*\|$	=	6.84336
$f(\underline{\alpha}^*, t_f^*)$	=	0.00199951
$e_x(\underline{\alpha}^*, t_f^*)$	=	$5.00358 \cdot 10^{-4}$
$e_v(\underline{\alpha}^*, t_f^*)$	=	$4.99823 \cdot 10^{-4}$
$\|\nabla f(\underline{\alpha}^*, t_f^*)\|$	=	$4.63842 \cdot 10^{-8}$
t_f^*	=	1.99901

Table 6.1: Training results for the car problem.

subject to the lower and upper bounds

$$a^*(t;\underline{\alpha}^*,t_f^*) = \begin{cases} 0, & a^*(t;\underline{\alpha}^*,t_f^*) < 0. \\ a^*(t;\underline{\alpha}^*,t_f^*), & 0 \leq a^*(t;\underline{\alpha}^*,t_f^*) \leq 1. \\ 1, & a^*(t;\underline{\alpha}^*,t_f^*) > 1. \end{cases} \tag{6.22}$$

$$d^*(t;\underline{\alpha}^*,t_f) = \begin{cases} -1, & d^*(t;\underline{\alpha}^*,t_f^*) < -1. \\ d^*(t;\underline{\alpha}^*,t_f^*), & -1 \leq d^*(t;\underline{\alpha}^*,t_f^*) \leq 0. \\ 0, & d^*(t;\underline{\alpha}^*,t_f^*) > 0. \end{cases} \tag{6.23}$$

and for $t \in [0, 1.99901]$.

The optimal control (acceleration and deceleration) and the corresponding optimal trajectories (position and velocity) obtained by the neural network are shown in Figures 6.5, 6.6, 6.7 and 6.8, respectively.

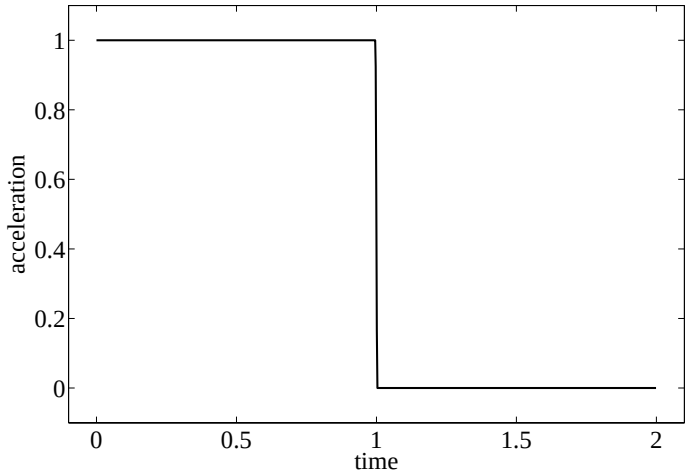


Figure 6.5: Neural network solution for the optimal acceleration in the car problem.

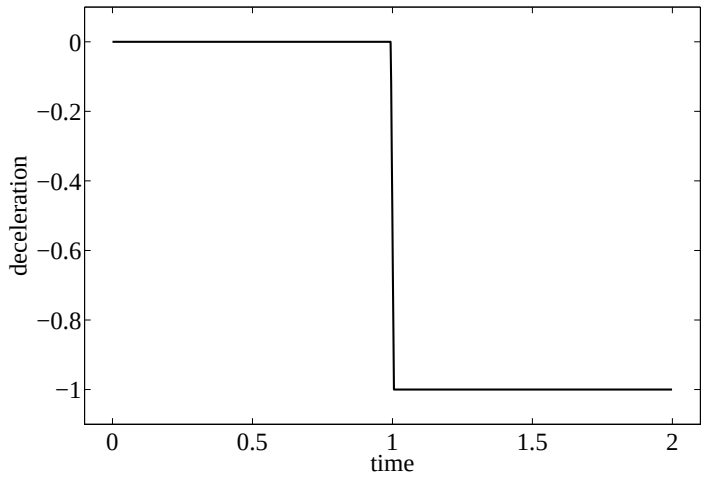


Figure 6.6: Neural network solution for the optimal deceleration in the car problem.

6.3 The fed batch fermenter problem

The fed batch fermenter problem for the multilayer perceptron is an optimal control problem with one control and four state variables, and defined by an objective functional with one constraint and requiring the integration of a system of ordinary differential equations. The implementation of this problem is included in Flood [61].

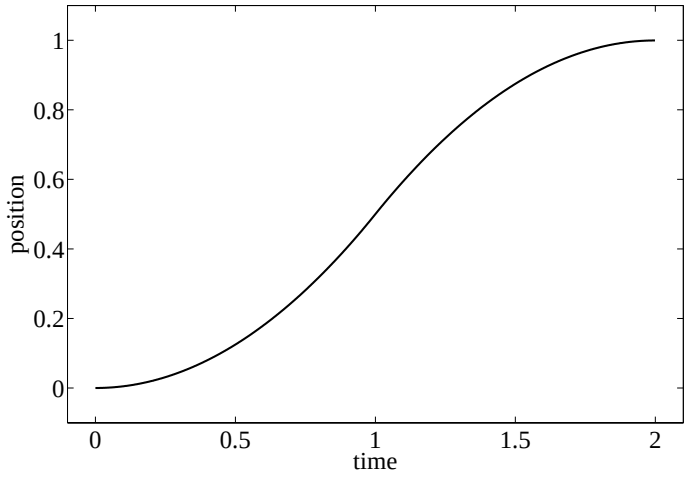


Figure 6.7: Corresponding optimal trajectory for the position in the car problem.

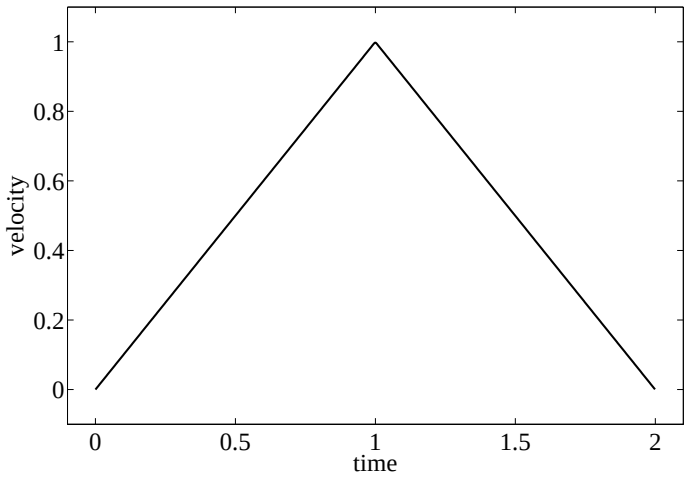


Figure 6.8: Corresponding optimal trajectory for the position in the car problem.

6.3.1 Introduction

In many biochemical processes, the reactors are operated in fed batch mode, where the feed rate into the reactor is used for control. There is no outflow, so the feed rate must be chosen so that that batch volume does not exceed the physical volume of the reactor.

As a specific example, an optimization study of the fed batch fermentation for ethanol production by Saccharomyces cerevisiae is presented.

In this Section, we seek to determine the optimal control law and the corresponding optimal trajectory of a fed batch fermenter problem using a neural network. We also compare the results

by this numerical method against those provided by other authors.

6.3.2 Problem statement

The fed batch fermentation process considered here is a process in which ethanol is produced by Saccharomyces cerevisiae and the production of ethanol is inhibited by itself.

A batch fermenter generally consists of a closed vessel provided with a means of stirring and with temperature control. It may be held at constant pressure or it can be entirely enclosed at a constant volume. In many biochemical processes, the reactors are operated in fed batch mode, where the feed rate into the reactor chosen so that that batch volume does not exceed the physical volume of the reactor [4]. Figure 6.3.2 shows the basics of the reactor.

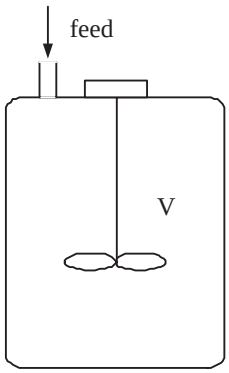


Figure 6.9: The fed batch fermenter.

State equations

The states of the plant are the concentration of cell mass $x(t)$, the concentration of substrate $s(t)$, the concentration of product $p(t)$ and the broth volume in the fermenter $v(t)$. The amount $u(t)$ is the feeding rate, which is the only manipulated variable of this process [23]. The dynamic behavior of this fed batch fermentation process can be described by the following differential-algebraic equations

$$\frac{dx(t)}{dt} = x(t)\mu(t) - u(t)\frac{x(t)}{v(t)}, \tag{6.24}$$

$$\frac{ds(t)}{dt} = -x(t)\frac{\mu(t)}{Y} + u(t)\frac{s_0 - s(t)}{v(t)}, \tag{6.25}$$

$$\frac{dp(t)}{dt} = -x(t)\eta(t) - u(t)\frac{p(t)}{v(t)}, \tag{6.26}$$

$$\frac{dv(t)}{dt} = u(t), \tag{6.27}$$

$$\mu(t) = \frac{\mu_0}{1 + \frac{p(t)}{K_p}}\frac{s(t)}{K_s + s(t)}, \tag{6.28}$$

$$\eta(t) = \frac{\nu_0}{1+\frac{p(t)}{K'_p}}\frac{s(t)}{K'_s+s(t)}, \tag{6.29}$$

together with their initial conditions

$$x(t_i) = x_i, \tag{6.30}$$
$$s(t_i) = s_i, \tag{6.31}$$
$$p(t_i) = p_i, \tag{6.32}$$
$$v(t_i) = v_i. \tag{6.33}$$

Here μ is the specific growth rate, η the specific productivity, Y the yield coefficient and s_0 the substrate concentration of the feed. The kinetic constants for Saccharomyces cerevisiae growing of glucose are $\mu_0 = 0.408\,h^{-1}$, $K_p = 16.0\,g\,l^{-1}$, $K_s = 0.22\,g\,l^{-1}$, $\eta_0 = 1.0\,h^{-1}$, $K'_p = 71.5\,g\,l^{-1}$ and $K'_s = 0.44\,g\,l^{-1}$ [46].

Input constraints

Here the feed rate to the reactor constrained to lie in the interval

$$u(t) \in [\inf(u), \sup(u)] \tag{6.34}$$

for $t \in [t_i, t_f]$.

State constraints

The liquid volume of the reactor is limited by the vessel size, $v(t) \leq V$. This constraint on the state of the system can be written as an error functional,

$$E_V[u(t)] \equiv V - v(t_f) = 0. \tag{6.35}$$

Performance requirements

The desired objective is to obtain a maximum amount of yield at the end of the process. The actual yield in the reactor is given by the concentration of product multiplied by the broth volume in the reactor. More specifically, the aim is to choose a feed rate which maximizes the functional

$$Y[u(t)] = p(t_f)v(t_f), \tag{6.36}$$

Since the equations describing the fermenter are nonlinear and the inputs and states are constrained, the determination of the feed rate to maximize the yield can be quite difficult.

Nominal values

The nominal values for all the parameters here are the same as those used by R. Luus in [70], so as to compare the neural network results to those reported by that author.

The yield coefficient, Y, is assumed to be a constant of 0.1. The initial state is specified as $x_0 = 1g\,l^{-1}$, $s_0 = 150g\,l^{-1}$, $p_0 = 0$ and $v_0 = 10\,l$, and the final time of the process, t_f, is set at $54\,h$. Besides, The feed rate to the reactor constrained by $0 \leq u(t) \leq 12$.

6.3.3 Numerical results

Selection of function space

The control variable $u(t)$ is to be represented by a multilayer perceptron with a hidden layer of sigmoid neurons and an output layer of linear neurons. Necessarily, the number of inputs is one, t, and the number of output neurons is also one, u. The size of the hidden layer is a design variable in the problem, which is set here to two. Figure 6.10 is a graphical representation of the resulting network architecture.

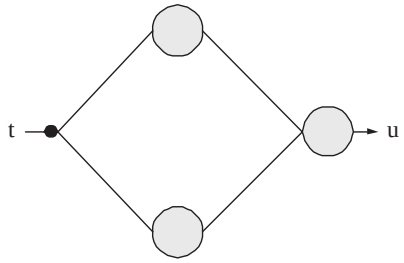


Figure 6.10: Network architecture for the fed batch fermenter problem.

This neural network spans a family V of parameterized functions $u(t;\underline{\alpha})$ of dimension $s = 7$, which is the number of biases and synaptic weights. Elements of V are of the form

$$\begin{array}{rccc} u: & \mathbb{R} & \to & \mathbb{R} \\ & t & \mapsto & u(t;\underline{\alpha}), \end{array}$$

where

$$u(t;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{2} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right). \tag{6.37}$$

The minimum and maximum input and output values are by far outside the range $[-1, 1]$. It is therefore necessary to pre and post-process input and output in order to achieve better approximation properties. The minimum and maximum values used here for pre and post-processing are listed in Table 6.2.

$\min(t)$	=	0
$\max(t)$	=	54
$\min(u(t;\underline{\alpha}))$	=	0
$\max(u(t;\underline{\alpha}))$	=	12

Table 6.2: Minimum and maximum for pre and post-processing in the fermenter problem.

In this way, the input is pre-processed to produce an input signal with minimum -1 and maximum 1,

$$t = 2\frac{t}{54} - 1. \tag{6.38}$$

Similarly, the output signal from the neural network is then post-processed to produce an output ranging from -1 and 1,

$$u(t;\underline{\alpha}) = 0.4\,(u(t;\underline{\alpha})12 + 1)\,. \tag{6.39}$$

It is easy to see that this form of pre and post-processing produces input and output signals in the range $[-1, 1]$.

On the other hand, the feed rate to the fermenter $u(t)$ is constrained to lie in the interval $[0, 12]$. To deal with such constraints we bound the network outputs in the form

$$u(t;\underline{\alpha}) = \begin{cases} 0, & u(t;\underline{\alpha}) < 0. \\ u(t;\underline{\alpha}), & 0 \leq u(t;\underline{\alpha}) \leq 12. \\ 12, & u(t;\underline{\alpha}) > 12. \end{cases} \tag{6.40}$$

Formulation of variational problem

Following Equations (6.35) and (6.36), the fed batch fermenter for the multilayer perceptron can be stated as to find an optimal control $u^*(t;\underline{\alpha}^*) \in V$ so that

$$E_V[u^*(t;\underline{\alpha}^*)] = 0, \tag{6.41}$$

and for which the functional

$$Y[u(t;\underline{\alpha})],$$

defined on V, takes on a minimum value.

This constrained problem can be formulated as an unconstrained one by adding a penalty term for the constraint to the original objective functional. More specifically, the unconstrained fed batch fermenter problem is stated as to find an optimal control $u^*(t;\underline{\alpha}^*) \in V$ for which the functional

$$F[u(t;\underline{\alpha})] = \rho_E(E_V[u(t;\underline{\alpha})])^2 - \rho_Y Y[u(t;\underline{\alpha})], \tag{6.42}$$

defined on V, takes on a minimum value.

The volume error an yield term weights ρ_E and ρ_Y, are set here to 10^{-3} and 10^{-10}, respectively.

Please note that evaluation of the objective functional (6.42) requires a numerical method for integration of ordinary differential equations. Here we choose the Runge-Kutta-Fehlberg method [93]. For this problem we set the tolerance to 10^{-12}.

On the other hand, there are no target outputs for the neural network here, so a back-propagation algorithm for the objective function gradient vector, $\nabla f(\underline{\alpha})$, can not be derived for this problem. Instead, we use the central differences method for numerical differentiation with $\epsilon = 10^{-6}$.

Solution of reduced function optimization problem

Here we use the quasi-Newton method for training. This method is applicable here, since the objective function is differentiable.

In this example, the training algorithm stops when it can not keep minimizing the objective function. When this situation occurs, the Brent's method gives zero train rate for a gradient descent train direction.

The training algorithm required 492 epochs to go. The evaluation of the initial guess is -0.0246886. After training this value falls to -0.0417673. On the other side, the gradient norm deceases from 0.0394488 to $2.7123 \cdot 10^{-6}$. Figures 6.11 and 6.12 explain this process in a graphical fashion.

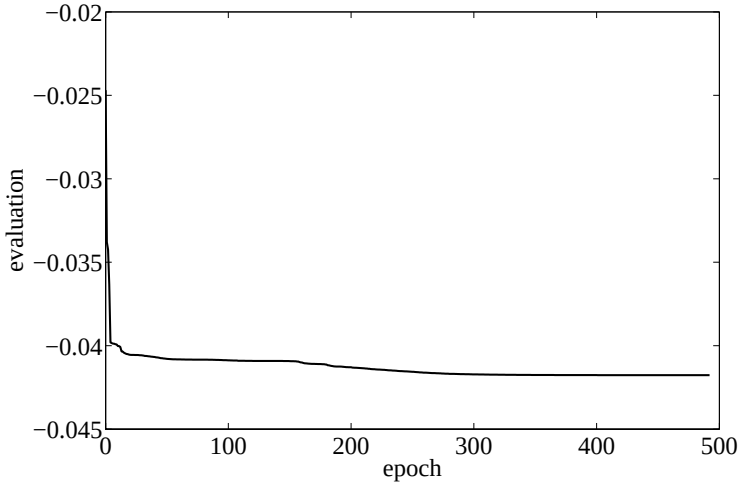


Figure 6.11: Evaluation history in the fed batch fermenter problem.

Table 6.3 shows the training results for this problem. Here N is the number of epochs, M the number of objective function evaluations, CPU the computing time in a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameters norm, $f(\underline{\alpha}^*)$ the final objective value, $e_v(\underline{\alpha}^*)$ the final error in the volume constraint and $y(\underline{\alpha}^*)$ the final yield. The final volume error is around 0.1% of the total volume, giving a very slight violation of the constraint. On the other side, the yield obtained by the neural network is about 0.2% higher than the 20406 l reported by Luus [70].

The optimal feed rate obtained by the neural network can be written in a explicit form as

$$\begin{aligned} t &= \frac{2}{54}t - 1, \\ u^*(t;\underline{\alpha}^*) &= -27.3844 \\ &+ 37.5292 \tanh\left(27.1799 - 26.3943t\right) \\ &+ 11.2443 \tanh\left(-1.83915 + 0.719688t\right), \\ u^*(t;\underline{\alpha}^*) &= 0.5(u^*(t;\underline{\alpha}^*) + 1)12, \end{aligned} \tag{6.43}$$

for $t \in [0, 54]$.

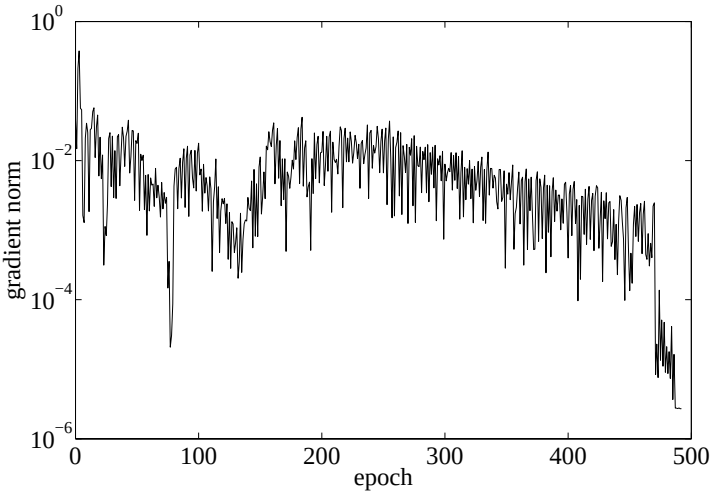


Figure 6.12: Gradient norm history in the fed batch fermenter problem.

N	=	493
M	=	24776
CPU	=	1019
$\|\underline{\alpha}^*\|$	=	23.0654
$f(\underline{\alpha}^*)$	=	$-4.17673 \cdot 10^{-2}$
$e_v(\underline{\alpha}^*)$	=	0.205
$y(\underline{\alpha}^*)$	=	20447.8
$\|\nabla f(\underline{\alpha}^*)\|$	=	$2.7123 \cdot 10^{-6}$

Table 6.3: Training results for the fed batch fermenter problem.

The optimal control obtained by the neural network is plotted in Figure 6.13. On the other hand, the optimal trajectories for the cell mass concentration, substrate concentration, product concentration and broth volume are depicted in Figures 6.14, 6.15, 6.16 and 6.17, respectively. All these process histories for the state variables are quite smooth functions. This is a desired property, since the process is desired to be stable. Finally, the optimal specific growth rate and productivity are plotted in Figures 6.18 and 6.19.

6.3.4 Conclusions

A variational formulation for neural networks provides a direct method for the solution of general variational problems. The method has been applied here to the fed batch fermenter optimal control problem.

One of the main advantages of this technique relies on the universal approximation property of the multilayer perceptron. Some other remarkable features are the simplicity of problem formulation and the no need for a good initial guess.

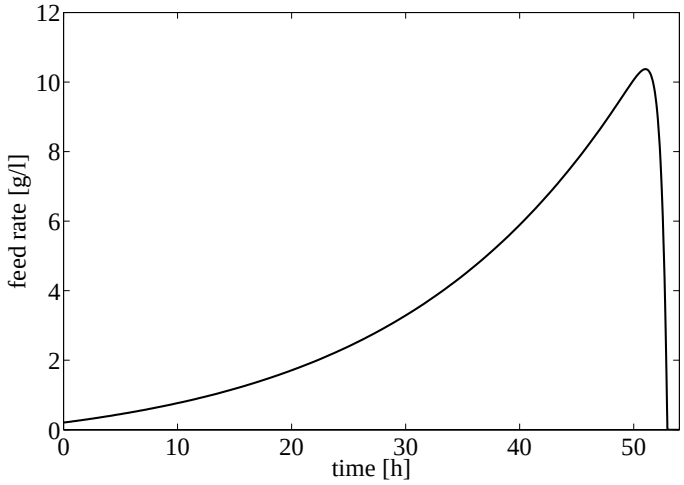


Figure 6.13: Optimal control for the fed batch fermenter.

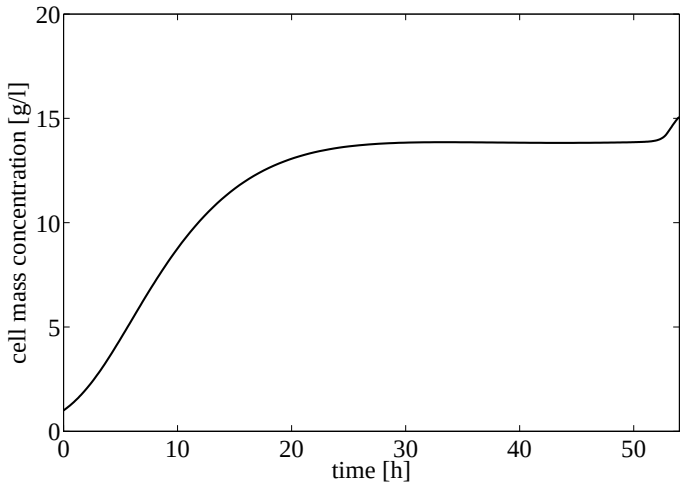


Figure 6.14: Optimal trajectory for the concentration of cell mass in fed batch fermenter.

6.4 The aircraft landing problem

This is an optimal control problem of aeronautical engineering interest, with one control and four state variables, and where the objective functional is evaluated by integrating a system of four ordinary differential equations. Also, the objective functional is built so as to find good performance in all the state variables. The source code for this application has been placed in the Flood library [61].

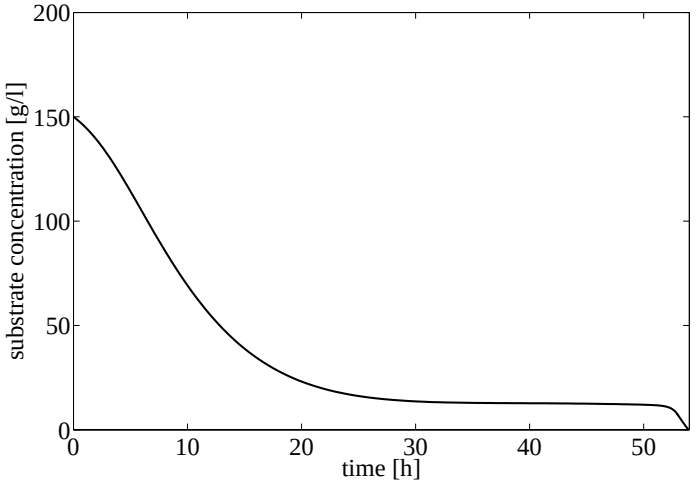


Figure 6.15: Optimal trajectory for the concentration of substrate in the fed batch fermenter.

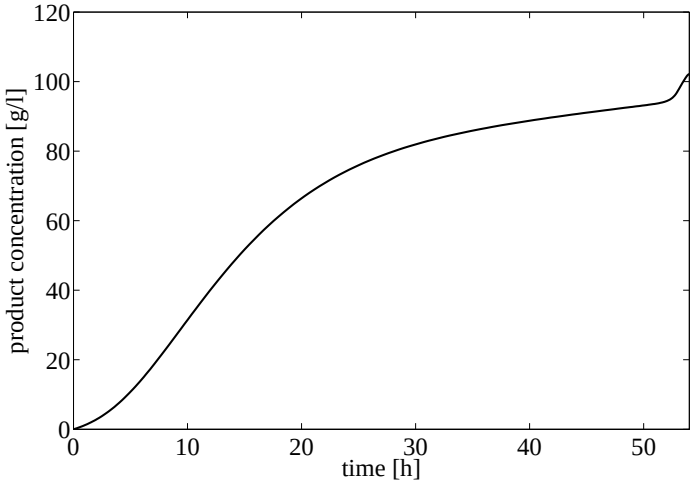


Figure 6.16: Optimal trajectory for the concentration of product in the fed batch fermenter.

6.4.1 Introduction

The landing of an aircraft consists of two main stages: the glide-path phase and the flare-out phase.

In the first stage the pilot is guided by an air traffic controller to a position where the aircraft will be within the range of the Instrument Landing System (ILS). At this position the pilot tunes the navigational receivers to the correct ILS frequency [1]. On board the aircraft a glide slope receiver utilizes this ILS signal to give the pilot a visual indication if the aircraft is currently above or below the desired glide path.

At approximately 30 meters above the runway the pilot begins the second and final stage of

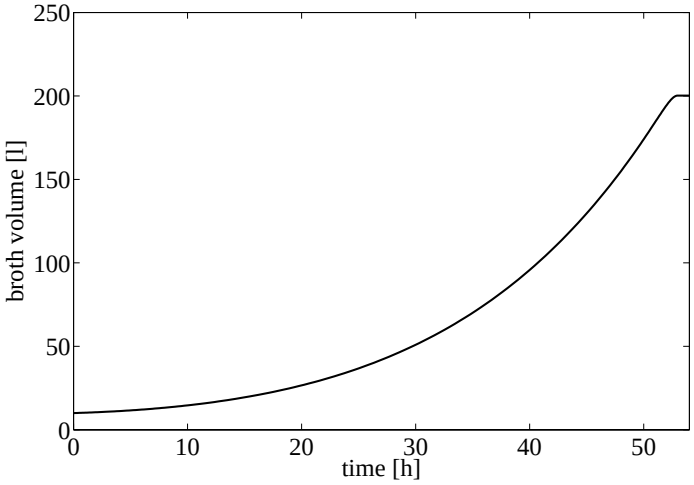


Figure 6.17: Optimal trajectory for the broth volume in the fed batch fermenter problem.

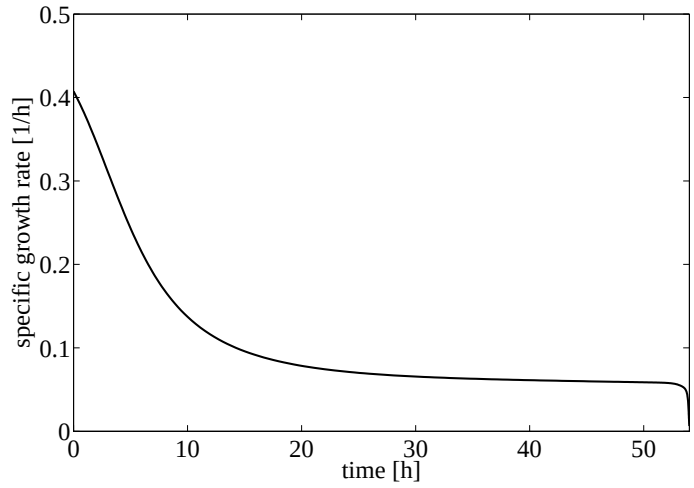


Figure 6.18: Optimal specific growth rate in the fed batch fermenter problem.

landing, the flare-out procedure. At this point the ILS glide slope becomes unsuitable and the pilot must guide the aircraft along the desired trajectory by making visual contact with the ground [28]. It is assumed that at the beginning of this flare-out phase the values of the altitude and the altitude rate lie within a given range of values, and that the aircraft has been guided to the correct position by air traffic controller. It is also assumed that during the flare-out the aircraft is not affected by any wind gust or perturbations.

Here we seek to determine the optimal control and the corresponding optimal trajectory of an aircraft during its final approach before landing. The aircraft landing problem examined here is

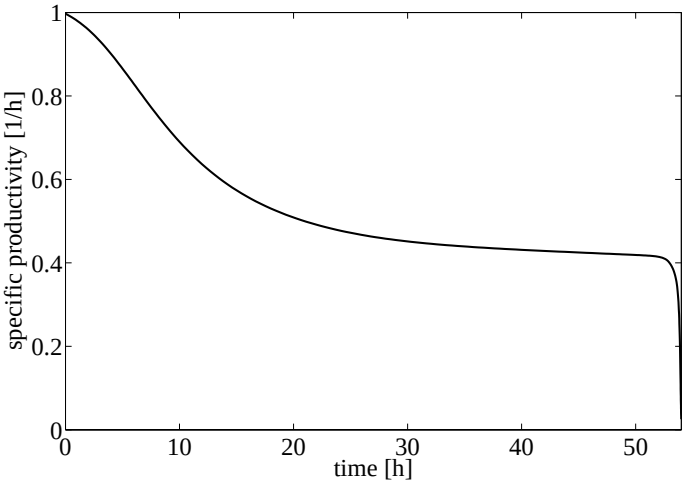


Figure 6.19: Optimal specific productivity in the fed batch fermenter problem.

similar to that considered in [28].

6.4.2 Problem statement

In this work our attention is focussed on the final phase of the landing process, the flare-out. This problem has been simplified as it is assumed that only the longitudinal motion of the aircraft need be considered, as the lateral motion is primarily used to set the orientation of the aircraft in the same direction as the runway, prior to the flare-out phase.

In the flare-out phase the longitudinal dynamics of the aircraft are controlled by the pitch angle. Changes in the pitch angle of are performed by varying the pitching moment around the center of mass. In civil aircrafts this is traditionally performed by a rotatable trailing edge flap situated on the horizontal stabilizer known as the elevator [5]. Thus the longitudinal motion of the aircraft is controlled by the magnitude of the elevator deflection angle. Figure 6.20 depicts the elevator deflection angle and the pitch angle of an aircraft.

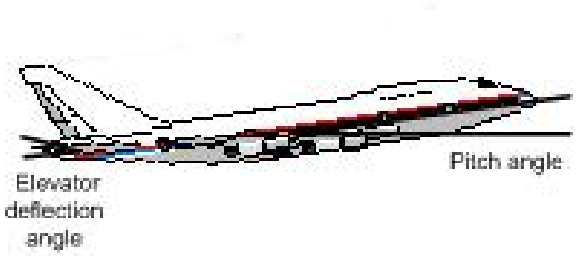


Figure 6.20: Elevator deflection angle and pitch angle.

Thus the aim of the aircraft landing problem considered here is to determine an optimal elevator deflection angle as a function of time that satisfies a set of performance requirements.

State equations

Using the simplified conditions stated above and the assumption that the glide angle, γ, is small ($\gamma \approx 0$), the equations of motion of the aircraft can be reduced to a set of equations known as the short period equations of motion [7].

These equations of motion can be written using a set of transfer functions relating the aircraft variables and the control input to the state variables, as shown in Equation (6.44). The variables of the aircraft motion used are the pitch angle rate (θ'), the altitude (h) and the altitude acceleration (h''). The control variable is the elevator deflection angle (δ).

The properties of the aircraft are defined using a set of parameters defined as the short period gain (K_s), the short period resonant frequency (ω_s), the short period damping factor (η) and the path time constant (T_s) [28]. The notation $'$ is used to denote the time derivative.

$$\begin{aligned}
\theta'(s) &= \frac{K_s(T_s s+1)}{\left(\frac{s^2}{\omega_s^2}+\frac{2\eta s}{\omega_s}+1\right)}\,\delta(s),\\
h''(s) &= \frac{V}{T_s s+1}\,\theta'(s),\\
h(s) &= \frac{1}{s^2}\,h''(s),\\
h(s) &= \frac{K_s V}{s^2\left(\frac{s^2}{\omega_s^2}+\frac{2\eta s}{\omega_s}+1\right)}\,\delta(s).
\end{aligned} \tag{6.44}$$

The variable V is the velocity of the aircraft and it is assumed to be constant during the flare-out phase at a value of $78\,ms^{-1}$. The aircraft parameters are also assumed to be time invariant, and the numerical values used here are $K_s = -0.95\,s^{-1}$, $T_s = 2.5\,s$, $\omega_s = -0.95\,rad\,s^{-1}$ and $\eta = 0.5$

However, some of the state variables used in Equation (6.44) are not readily available. For example the term h'' can be difficult to obtain. For this reason the following variables will be used instead to describe the dynamics of the aircraft [28]: θ', the pitch angle rate, θ, the pitch angle, h', the altitude rate and h, the altitude. This set of variables can be easily obtained from gyroscopic or radar altimeter measurements in flight. Equation (6.44) can then be transformed into the following set of ordinary differential equations in the time domain

$$\begin{aligned}
\frac{d\theta'}{dt} &= \left(\frac{1}{T_s}-2\eta\omega_s\right)\theta'(t)\\
&+ \left(\frac{2\eta\omega_s}{T_s}-\omega_s^2-\frac{1}{T_s^2}\right)\theta(t)\\
&+ \left(\frac{1}{VT_s^2}-\frac{2\eta\omega_s}{VT_s}+\frac{\omega_s^2}{V}\right)h'(t)\\
&+ \omega_s^2 K_s T_s \delta(t),\\
\frac{d\theta}{dt} &= \theta'(t),\\
\frac{dh'}{dt} &= \frac{V}{T_s}\theta(t)-\frac{1}{T_s}h'(t),\\
\frac{dh}{dt} &= h'(t).
\end{aligned} \tag{6.45}$$

It can be seen here that the elevator deflection angle (δ) has a direct effect on the pitch angle rate (θ'), which in turn affects the pitch angle (θ), the altitude rate (h') and the altitude (h).

Performance requirements

The performance requirements define the physical constraints and desired values of the control and the state variables. The most important requirements and constraints for the landing system considered in this problem are highlighted in the following section.

In our example problem the flare-out procedure begins at $t_i = 0s$ and ends at the final or touchdown time $t_f = 20s$. The initial conditions of the state variables at t_i are displayed in Table 6.4.

h_0	$30\,m$
h'_0	$6\,m\,s^{-1}$
θ_0	$-0.078\,rad$
θ'_0	$0\,rad\,s^{-1}$

Table 6.4: Initial Conditions for the aircraft landing problem.

During a process it is often desirable to be able to define the desired value of a given state variable; this information can then be used to evaluate the performance of the system. The desired altitude of the aircraft is the most visual characteristic of the landing procedure. For this problem it is given by the following expression

$$h_d(t) = \begin{cases} 30\exp\left(-\frac{t}{5}\right), & 0 \leq t \leq 15, \\ 6 - 0.3t, & 15 \leq t \leq 20. \end{cases} \tag{6.46}$$

This exponential-linear trajectory ensures a safe and comfortable landing. The desired altitude is displayed in Figure 6.21.

An important requirement is that the aircraft does not land before the final touchdown time, t_f. At this time the following condition must be satisfied

$$h_d(t_f) = 0. \tag{6.47}$$

The desired altitude rate, or vertical velocity, of the aircraft is given by the time derivative of the desired altitude expression,

$$h'_d(t) = \begin{cases} 6\exp\left(-\frac{t}{5}\right), & 0 \leq t \leq 15, \\ -0.3, & 15 \leq t \leq 20. \end{cases} \tag{6.48}$$

At the time of touchdown the pitch angle of the aircraft must lie in the range $[0°, 10°]$. This requirement is defined by a set of physical limitations. The lower limit serves to ensure the nose wheel of a tricycle landing gear does not touchdown prematurely. Similarly the upper limit is set to prevent the tail gear touching downing first. A desired pitch angle at touchdown could be specified as

$$\theta_d(t_f) = 2°. \tag{6.49}$$

In order to ensure safety and comfortability during the landing phase, it is desirable to restrict the pitch angle rate from excessive fluctuations. Thus the desired pitch angle rate is

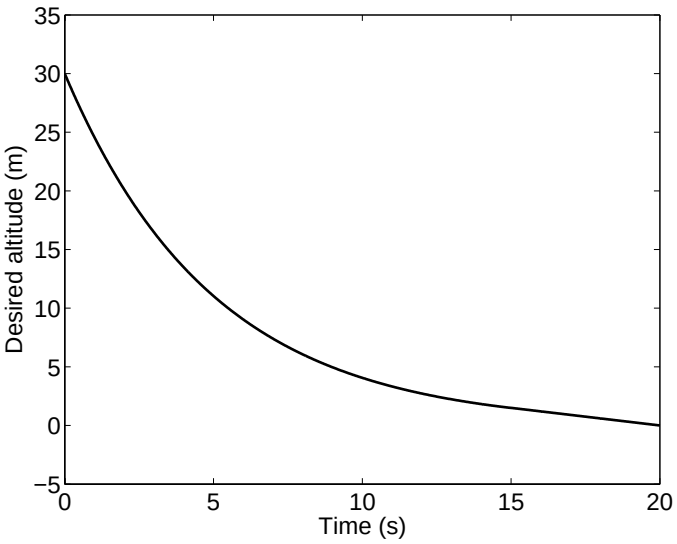


Figure 6.21: Desired altitude in the aircraft landing problem.

$$\theta'_d(t) = 0, \quad 0 \leq t \leq t_f. \tag{6.50}$$

As stated earlier, the elevator controls the longitudinal motion of the aircraft. It is assumed that any control signal is instantaneously represented by the elevator. The elevator deflection angle is also physically limited to a finite range

$$-35^\circ \leq \delta(t) \leq +15^\circ, \quad 0 \leq t \leq t_f. \tag{6.51}$$

It is desirable to land without expending excessive amounts of control effort. Therefore the desired elevator deflection angle is defined as

$$\delta_d(t) = 0, \quad 0 \leq t \leq t_f. \tag{6.52}$$

6.4.3 Numerical results

In this Section a multilayer perceptron is trained to find the optimal control and the corresponding optimal trajectory for the aircraft landing problem formulated in Section 6.4. The problem is solved using the Flood library [61].

Selection of function space

The first step in solving this problem is to choose a network architecture, in order to define a function space for the control variable. Here a multilayer perceptron with a sigmoid hidden layer and a linear output layer is used. This neural network is a class of universal approximator [49]. The neural network has one input, the time, and one output neuron representing the elevator deflection angle. We take an initial guess at the number of neurons in the hidden layer to be three. This neural network can be denoted as a $1:3:1$ multilayer perceptron. Figure 6.22 is a graphical representation of this network architecture.

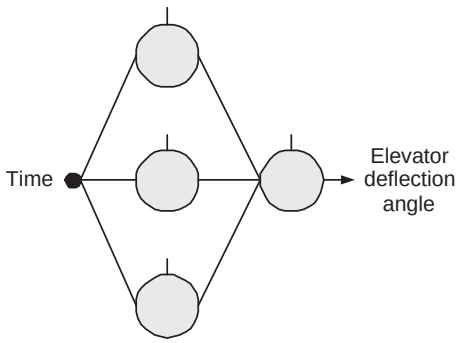


Figure 6.22: Network architecture for the aircraft landing problem.

Such a multilayer perceptron spans a family V of parameterized functions $\delta(t;\underline{\alpha})$ of dimension $s = 10$, which is the number of parameters in the neural network network. Elements V are of the form

$$\begin{array}{rccc} \delta: & \mathbf{R} & \rightarrow & \mathbf{R} \\ & t & \mapsto & \delta(t;\underline{\alpha}), \end{array}$$

where

$$\delta(t;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \cdot \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right). \tag{6.53}$$

The parameters of the neural network are initialized so that the elevator deflection angle $\delta(t;\underline{\alpha})$ is 0 for all landing time $0 \leq t \leq t_f$.

The elevator deflection angle is constrained to lie in the range $[-35°, +15°]$. The output of the neural network is bounded in the following way

$$\delta(t;\underline{\alpha}) \quad = \quad \begin{cases} -35°, & \delta(t;\underline{\alpha}) < -35°. \\ \delta(t;\underline{\alpha}), & -35° \leq \delta(t;\underline{\alpha}) \leq +15°. \\ +15°, & \delta(t;\underline{\alpha}) > +15°. \end{cases} \tag{6.54}$$

Formulation of variational problem

The second step is to select a suitable objective functional in order to formulate the variational problem. This functional will determine the form of the optimal control function ($\delta^*(t)$) and is based upon the performance requirements discussed in Section 6.4.2.

From Equations (6.46) (6.47), (6.48), (6.49), (6.50) and (6.52) the objective functional used in this problem is defined as

$$\begin{array}{rcl} F[\delta(t;\underline{\alpha})] & = & \alpha_h \displaystyle\int_0^{t_f} [h(t) - h_d(t)]^2 dt \\ & + & \beta_h h_d(t_f) \end{array}$$

$$
\begin{aligned}
&+ \quad \alpha_{h'} \int_0^{t_f} [h'(t) - h'_d(t)]^2 dt \\
&+ \quad \beta_\theta \theta_d(t_f) \\
&+ \quad \alpha_{\theta'} \int_0^{t_f} [\theta'(t) - \theta'_d(t)]^2 dt \\
&+ \quad \alpha_\delta \int_0^{t_f} [\delta(t) - \delta_d(t)]^2 dt,
\end{aligned} \tag{6.55}
$$

where α_h, $\alpha_{h'}$, $\alpha_{\theta'}$ and α_δ are the altitude, altitude rate, pitch angle rate and elevator deflection angle weight factors; the terms β_h and β_θ are the touchdown weight factors for the altitude and the pitch angle. Table 6.5 displays the set used in this investigation. These numbers are design variables of the problem and were obtained with some trial and error.

α_h	5.0×10^1
β_h	5.0×10^3
$\alpha_{h'}$	1.0×10^{-4}
β_θ	2.0×10^6
$\alpha_{\theta'}$	2.0×10^6
α_δ	1.0×10^5

Table 6.5: Weight factors in the aircraft landing problem.

Note that evaluation of the objective functional, Equation (6.55), requires the time history of all the state variables in response to the time history of the control variable input. This is determined by integrating the state equations of the system, numerical integration is performed here using the Runge-Kutta method with 1000 integration points [83]. Note also that this objective functional also requires a numerical method for integration of functions in order to be evaluated; in this problem we use the trapezoid method.

Solution of reduced function optimization problem

The third step is to choose a suitable training algorithm, in order to solve the reduced function optimization problem. Here we use a quasi-Newton method with BFGS train direction and Brent optimal train rate methods [83]. The tolerance in the Brent's method is set to 10^{-6}.

The objective function gradient vector $\nabla f(\underline{\alpha})$ must be evaluated via numerical differentiation. Here we use the symmetrical central differences method [15] with an epsilon value of 10^{-6}.

The training algorithm is set to stop when the optimal train rate in Brent's method is 0. For this case, the quasi-Newton method required 291 epochs or iterations to find the minimum of the objective functional. The evaluation of the initial guess was 4.76; after training this value falls to 1.002×10^{-4}. Figure 6.23 displays the training history for this problem, showing objective functional evaluation against the number of training epochs. Note that the Y-axis uses a logarithmic (base 10) scale.

Table 6.6 shows the training results for this problem. Here N denotes the number of epochs, M the number of objective function evaluations, $f(\underline{\alpha}^*)$ the final objective function value and $\|\nabla f(\underline{\alpha}^*)\|$ the final objective function gradient norm. From this table it can be seen that the final gradient norm approaches a very small value, which indicates that the training algorithm has converged to a minimum point.

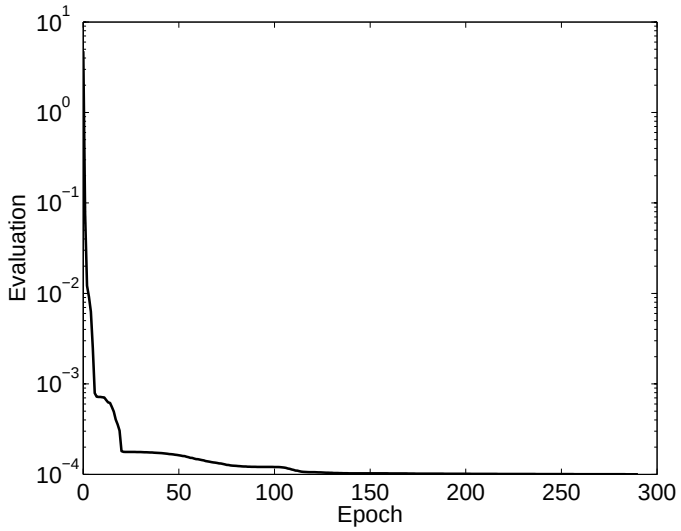


Figure 6.23: Evaluation history for the aircraft landing problem.

N	291
M	16748
$f(\underline{\alpha}^*)$	1.002×10^{-4}
$\|\nabla f(\underline{\alpha}^*)\|$	6.09×10^{-2}

Table 6.6: Training results for the aircraft landing problem.

The optimal control obtained by the neural network is shown in Figure 6.24. The corresponding optimal trajectories of the altitude, altitude rate, pitch angle and pitch angle rate are shown in Figures 6.25, 6.26, 6.27 and 6.28 respectively.

Figure 6.24 shows the elevator deflection angle during the landing phase. It can be seen that the magnitude of control input ranges from -3° to 0° and has a smooth profile. This therefore satisfies the requirement of reducing the required work effort.

The optimal trajectory for the altitude, depicted in Figure 6.25, matches the desired altitude profile detailed in Equation (6.46). It can be seen that at the final touchdown time, t_f, the altitude of the aircraft is $0\,m$.

Figure 6.26 shows that the altitude rate of the aircraft; it can be seen that it follows the profile specified by Equation (6.48).

The pitch angle variation during the landing phase is displayed in Figure 6.27. The value at touchdown is approximately $0.5°/s$, this value lies within the desired range $[0°, 10°]$. However the desired touchdown value of 2° has not been obtained. This is believed to be a result of the simplified state equations used; for example the model does not account for effects such as induced lift at low altitudes.

Finally Figure 6.28 displays the pitch angle rate. It can be seen that throughout the landing phase the magnitude of the pitch angle rate is relatively small and the profile is sufficiently smooth to ensure a comfortable and safe landing.

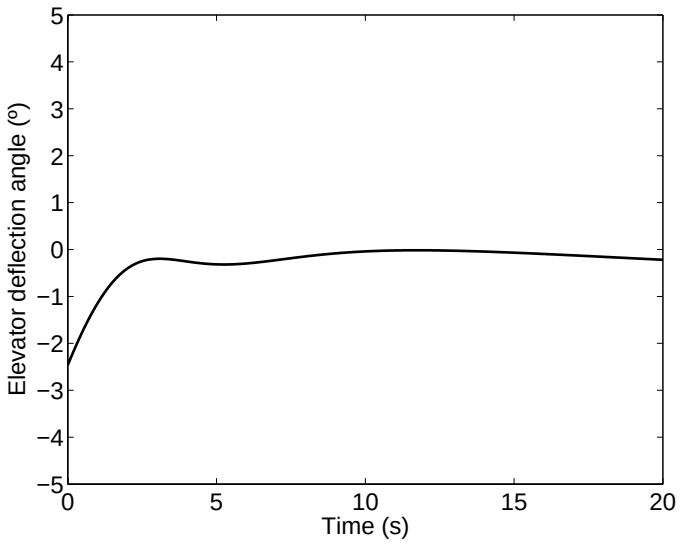


Figure 6.24: Optimal control (elevator deflection angle) for the aircraft landing problem.

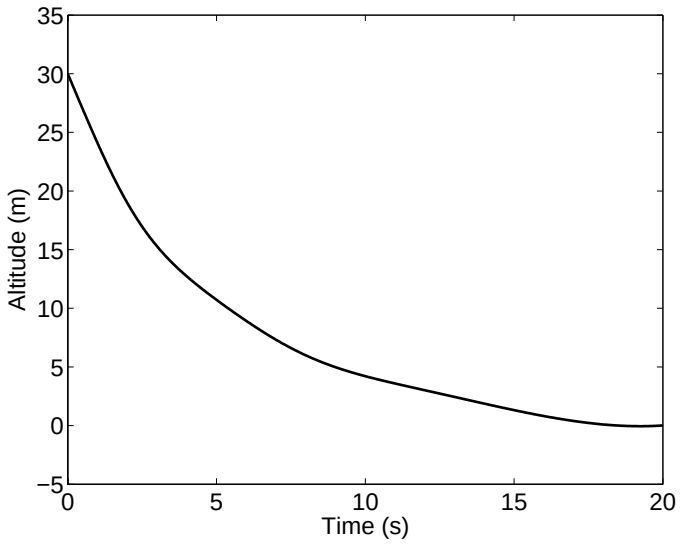


Figure 6.25: Optimal altitude trajectory for the aircraft landing problem.

6.4.4 Conclusions

In this work the variational formulation for the multilayer perceptron has been applied to find the optimal control and the corresponding optimal trajectory of an aircraft landing system. The results of the neural network are shown to give good performance qualities for the control inputs and the state variables of the system.

However, this application does need some refinement. In particular, a more complete set of

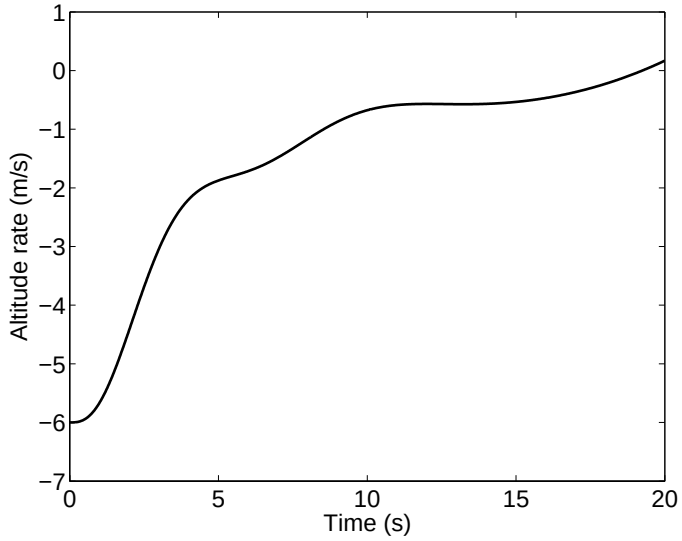


Figure 6.26: Optimal altitude rate trajectory for the aircraft landing problem.

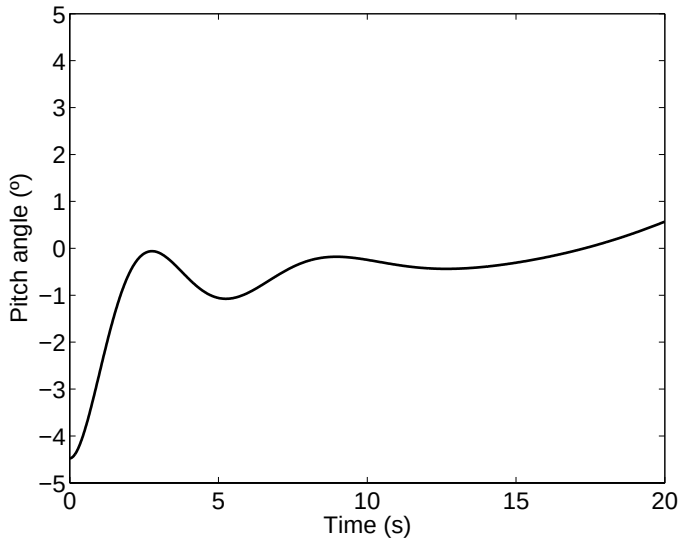


Figure 6.27: Optimal pitch angle trajectory for the aircraft landing problem.

state equations describing the aircraft dynamics should be used.

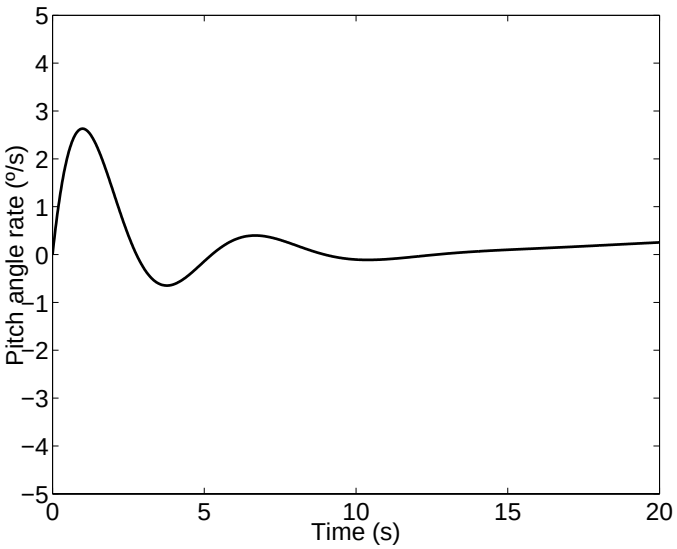


Figure 6.28: Optimal pitch angle rate trajectory for the aircraft landing problem.

This page is intentionally left blank.

Chapter 7

Inverse problems

Here the main issues concerning the theory of inverse problems are covered. These concern variational nature and ill-posseness, so that neural networks can be trained to learn the solution of this type of problems.

Two artificially generated inverse problems are here attempted for validation, achieving good results. As they make use of the finite element method, neural networks can be seen as a valuable tool for real problems. A real application in the metallurgical industry is also satisfactorily resolved in this section.

7.1 Problem formulation

Inverse problems can be described as being opposed to direct problems. In a direct problem the cause is given, and the effect is determined. In an inverse problem the effect is given, and the cause is estimated [54]. There are two main types of inverse problems: input estimation, in which the system properties and output are known and the input is to be estimated; and properties estimation, in which the the system input and output are known and the properties are to be estimated [54]. Inverse problems are found in many areas of science and engineering.

An inverse problem is specified by:

(i) A mathematical model of the system.

(ii) A set of actual observations to that system.

(iii) A statement of the physical constraints.

(iv) A specification of the error measure.

Mathematical model

The mathematical model can be defined as a representation of the essential aspects of some system which presents knowledge of that system in usable form.

Let us represent $\mathbf{y}(\mathbf{x})$ the vector of unknowns (inputs or properties) and $\mathbf{u}(\mathbf{x})$ the vector of state variables. The mathematical model, or state equation, relating unknown and state variables takes the form

$$\mathcal{L}(\mathbf{y}(\mathbf{x}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}, \tag{7.1}$$

where $\mathcal{L}$ is some algebraic or differential operator and $\mathbf{f}$ is some forcing term.

Observed data

Inverse problems are those where a set of measured results is analyzed in order to get as much information as possible on a mathematical model which is proposed to represent a real system.

Therefore, a set of experimental values on the state variables is needed in order to estimate the unknown variables of that system. This observed data is here denoted $\widehat{\mathbf{u}}(\mathbf{x})$.

In general, the observed data is invariably affected by noise and uncertainty, which will translate into uncertainties in the system inputs or properties.

Physical constraints

For inverse problems, the presence of restrictions is typical. Two possible classes of constraints are unknowns and state constraints.

The former are defined by the allowable values on the inputs or the properties of the system, depending on whether the we are talking about an input or a property estimation problem. Two typical types of constraints here are boundary conditions and lower and upper bounds.

State constraints are those conditions that the system needs to hold. This type of restrictions depend on the particular problem.

In this way, an unknown which satisfies all the input and state constraints is called an admissible unknown [53].

Definition 12 (Admissible unknown)**.** An unknown $\mathbf{y}(\mathbf{x})$ which satisfies all the constraints is called an admissible unknown. The set of admissible unknowns can be denoted Y, and the notation $\mathbf{y}(\mathbf{x}) \in Y$ means that the unknown $\mathbf{y}(\mathbf{x})$ is admissible.

Also, a state which satisfies the state constraints is called an admissible state [53].

Definition 13 (Admissible state)**.** A state $\mathbf{u}(\mathbf{x})$ which satisfies the state constraints is called an admissible state. The set of admissible states will be denoted by $\mathbf{U}$, and $\mathbf{u}(\mathbf{x}) \in \mathbf{U}$ means that $\mathbf{u}(\mathbf{x})$ is admissible.

Error functional

The inverse problem provides a link between the outputs from the model and the observed data. When formulating and solving inverse problems the concept of error functional is used to specify the proximity of the state variable $\mathbf{u}(\mathbf{x})$ to the observed data $\widehat{\mathbf{u}}(\mathbf{x})$:

$$\begin{array}{rccc} E: & Y & \to & \mathbb{R} \\ & \mathbf{y}(\mathbf{x}) & \mapsto & E[\mathbf{y}(\mathbf{x})]. \end{array}$$

The error functional $E[\mathbf{y}(\mathbf{x})]$ is of the form

$$E[\mathbf{y}(\mathbf{x})] = \|\mathbf{u}(\mathbf{x}) - \widehat{\mathbf{u}}(\mathbf{x})\|, \tag{7.2}$$

where any of the generally used norms may be applied to characterize the proximity of $\mathbf{u}(\mathbf{x})$ and $\widehat{\mathbf{u}}(\mathbf{x})$. Some of them are the sum squared error or the Minkowski error. Regularization theory can also be applied here [18].

The solution of inverse problems is then reduced to finding of extremum of a functional:

Problem 9 (Inverse problem)**.** Let Y and U be the function spaces of all admissible unknowns and states, respectively. Find an admissible unknown $\mathbf{y}^*(\mathbf{x}) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x};\underline{\alpha}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to follow an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$, and for which the error functional

$$E[\mathbf{y}(\mathbf{x})],$$

defined on Y, takes on a minimum value.

On the other hand, inverse problems might be ill-posed [95]. A problem is said to be well possed if the following conditions are true: (a) the solution to the problem exists ; (b) the solution is unique; and (c) the solution is stable. This implies that for the above-considered problems, these conditions can be violated. Therefore, their solution requires application of special methods. In this regard, the use of regularization theory is widely used [30].

Solution approach

In some elementary cases, it is possible to establish analytic connections between the sought inputs or properties and the observed data. But for the majority of cases the search of extrema for the error functional must be carried out numerically, and the so-called direct methods can be applied.

A variational formulation for neural networks provides a direct method for the solution of variational problems, and therefore inverse problems. Following the nomenclature here introduced, the inverse problem for the multilayer perceptron can be formulated as follows:

Problem 10 (Inverse problem for the multilayer perceptron)**.** Let Y and U be the function spaces of admissible unknowns and states, respectively. Let also V be the space consisting of all unknowns $\mathbf{y}(\mathbf{x};\underline{\alpha})$ that a given multilayer perceptron can define, with dimension s. Find an admissible unknown $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x};\underline{\alpha}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to be in an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$, and for which the error functional

$$E[\mathbf{y}(\mathbf{x};\underline{\alpha})],$$

defined on V, takes on a minimum value.

7.2 Validation examples

A variational formulation for the multilayer perceptron is general enough to incorporate a large number of novel applications such as inverse problems. This class of problems is illustrated with several examples for validation. The selected applications are easy to understand and are artificially generated, thus allowing to compare the neural network results against the exact ones.

7.2.1 The boundary temperature estimation problem

This is an input estimation problem, with one input and one output. The error functional is computed by integrating a partial differential equation. This is done with the finite element method.

Problem statement

Consider the heat transfer equation in a domain Ω with boundary Γ,

$$\nabla^2 T(x,y;t) = \frac{\partial T(x,y;t)}{\partial t} \quad in \quad \Omega, \tag{7.3}$$

$$T(x,y;0) = T_0 \quad in \quad \Omega, \tag{7.4}$$

$$T(x,y;t) = T_\Gamma(t) \quad on \quad \Gamma, \tag{7.5}$$

for $t \in [0, t_f]$, and where T_0 is the initial temperature and $T_\Gamma(t)$ is the boundary temperature, which varies with time.

On the other hand, consider a set of temperature measurements at some point (a, b) on Ω,

$$\begin{array}{cc} t_1 & \hat{T}_1 \\ t_2 & \hat{T}_2 \\ \vdots & \vdots \\ t_P & \hat{T}_P \end{array}$$

where P is the number of observed samples.

The problem is to estimate the boundary temperature $T_\Gamma(t)$ for $t \in [0, t_f]$, from the heat transfer model and the observed temperature data. Figure 7.1 is a graphical statement of the input estimation problem considered here.

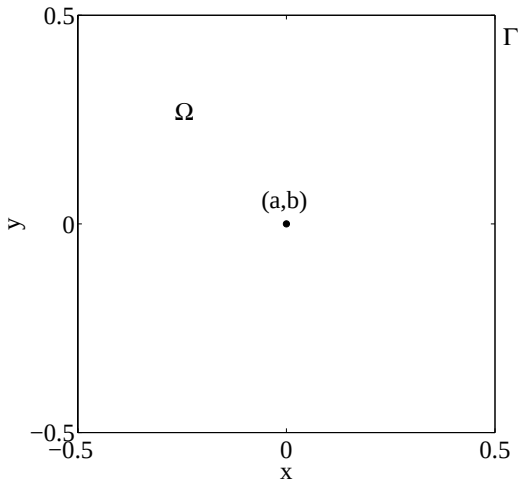


Figure 7.1: The boundary temperature estimation problem statement.

The selected error functional for this inverse problem is the mean squared error between the model temperatures and the experimental temperatures. The argument of this functional is the boundary temperature which is to be found,

$$E[T_\Gamma(t)] = \frac{1}{P}\sum_{i=1}^{P}\left(T(a,b;t_i) - \hat{T}_i\right)^2. \tag{7.6}$$

In this case study the domain is taken to be the square $\Omega = \{(x,y) : |x| \leq 0.5, |y| \leq 0.5\}$ with boundary $\Gamma = \{(x,y) : |x| = 0.5, |y| = 0.5\}$.

A set of 101 temperature samples at the center of the square $(a,b) = (0,0)$ is artificially generated from the following expression for the boundary temperature, which is therefore considered to be the analytical solution to the problem,

$$T_\Gamma(t) = \frac{\exp\left(-(t-\mu)^2/2\sigma^2\right)}{\sigma\sqrt{2\pi}/8} \tag{7.7}$$

for $t = [0,1]$ and with $\mu = 0.5$ and $\sigma = 0.05$. Figure 7.2 shows the analytical boundary temperature in Equation (7.7) and its corresponding center temperature.

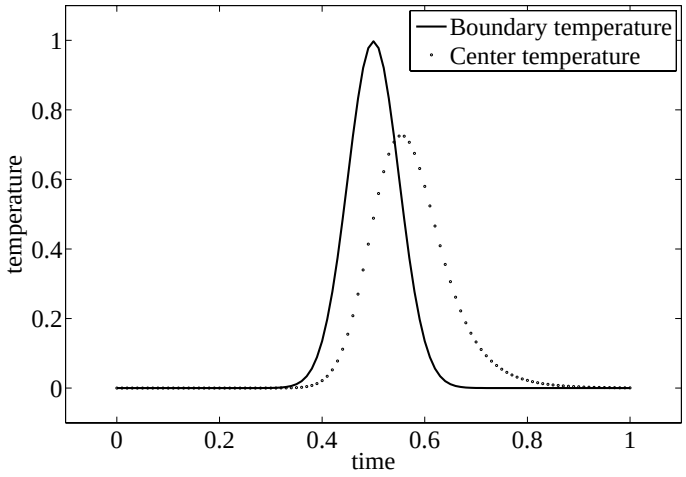


Figure 7.2: Analytical boundary temperature and corresponding center temperature.

Selection of function space

The neural network used to represent the boundary temperature $T_\Gamma(t)$, for $t \in [0,1]$, is a multilayer perceptron with 1 input, 3 sigmoid hidden neurons and 1 linear output neuron. It constructs a family V of parameterized functions $T_\Gamma(t;\underline{\alpha})$ of dimension $s = 10$, which is the total number of biases and synaptic weights. Figure 7.3 is a graphical representation of this network architecture.

Elements of V are written

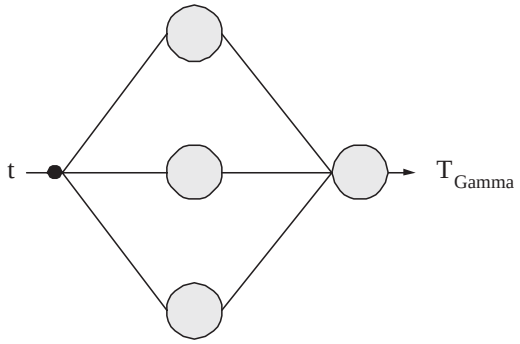


Figure 7.3: Network architecture for the boundary temperature estimation problem.

$$\begin{array}{rccc} y: & \mathbb{R} & \rightarrow & \mathbb{R} \\ & x & \mapsto & y(x;\underline{\alpha}), \end{array}$$

with

$$T_\Gamma(t;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} t\right). \tag{7.8}$$

The free parameters are here initialized at random. This provides a random initial boundary temperature.

Formulation of variational problem

From Equation (7.6) the statement of this problem is to find a function $T_\Gamma^*(t;\underline{\alpha}^*) \in V$ for which the error functional

$$E[T_\Gamma(t;\underline{\alpha})] = \frac{1}{P}\sum_{i=1}^{P}\left(T(a,b;t_i) - \hat{T}_i\right)^2, \tag{7.9}$$

defined on V, takes on a minimum value.

Evaluation of the objective functional (7.9) requires a numerical method for integration of partial differential equations. Here we choose the finite element method [102]. For this problem we use a triangular mesh with 888 elements and 485 nodes.

The objective function gradient vector $\nabla f(\underline{\alpha})$ [15] is carried out by means of numerical differentiation. In particular, we use the symmetrical central differences method [15] with an epsilon value of 10^{-6}.

Solution of reduced function optimization problem

Here we use a quasi-Newton method with BFGS train direction and Brent optimal train rate methods for training [15]. The tolerance in the Brent's method is set to 10^{-6}. While other

methods might suffer from local optima in this problem, the neural networks method proposed here has demonstrated convergence to the global optimum in this problem.

In this example, we set the training algorithm to stop when the optimal training rate for a gradient descent direction is zero. That means that the necessary condition for a local minimum has about been satisfied.

The initial mean squared error is 0.443 and the gradient norm 0.230. During the training process the objective function decreases until the stopping criterium is satisfied. This happens at epoch 95, when the mean squared error has fallen to $5.117 \cdot 10^{-5}$ and the gradient norm is $1.252 \cdot 10^{-3}$. Figures 7.4 and 7.5 show this training history for the evaluation and the gradient norm variables, respectively. The y-axis in in logarithmic scale at both plots.

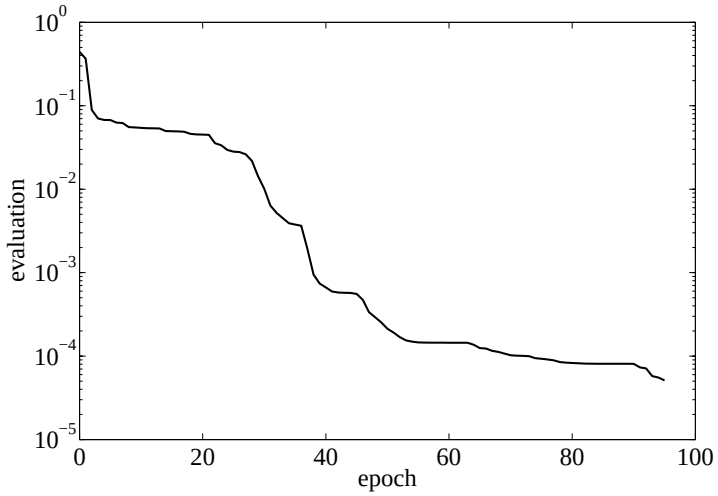


Figure 7.4: Evaluation history for the boundary temperature estimation problem.

Table 7.1 shows the training results for this problem. Here N denotes the number of epochs, M the number of objective evaluations, CPU the computing time in a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameters norm, $e(\underline{\alpha}^*)$ the final mean squared error value, and $\|\nabla e(\underline{\alpha}^*)\|$ the final objective function gradient norm.

N	=	95
M	=	7845
CPU	=	03 : 23 : 09
$\|\underline{\alpha}^*\|$	=	24.423
$e(\underline{\alpha}^*)$	=	$5.117 \cdot 10^{-5}$
$\|\nabla e(\underline{\alpha}^*)\|$	=	$1.252 \cdot 10^{-3}$

Table 7.1: Training results for the boundary temperature estimation problem.

Figure 7.6 shows the boundary temperature estimated by the neural network and the corresponding temperature at the center of the square. Comparing Figures 7.2 and 7.6 we can see that the estimated boundary temperature matches the actual boundary temperature to a great extent.

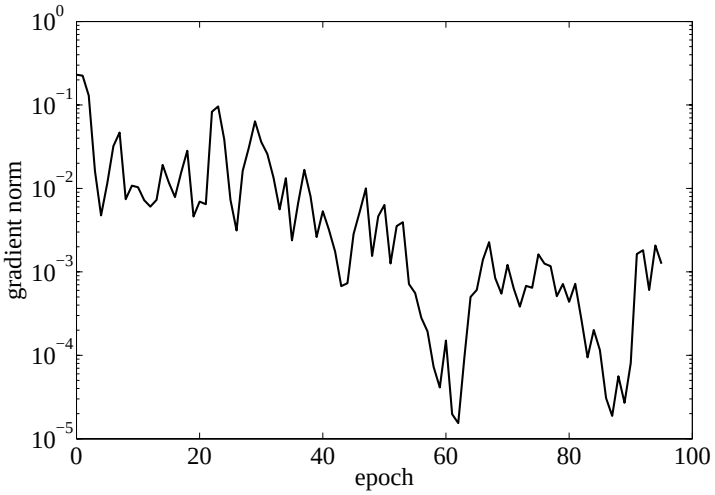


Figure 7.5: Gradient norm history for the boundary temperature estimation problem.

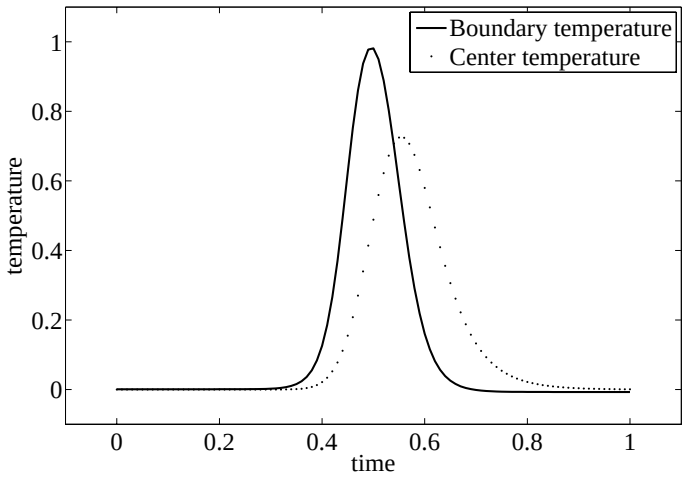


Figure 7.6: Neural network results for the boundary and corresponding center temperatures.

The expression for the boundary temperature by the trained neural network is as follows,

$$\begin{aligned} T_{\Gamma}^{*}(t;\underline{\alpha}^{*}) &= 0.101729 \\ &- 0.681005\tanh\left(-9.55196 + 17.5454t\right) \\ &- 0.104851\tanh\left(11.6386 + 69.6148t\right) \\ &- 0.677045\tanh\left(9.90625 + -21.9968t\right). \end{aligned} \tag{7.10}$$

7.2.2 The thermal conductivity estimation problem

Problem statement

For this problem, consider a 2-dimensional inhomogeneous medium with domain Ω and boundary Γ, in which the thermal conductivity is to be estimated. The mathematical model of heat transfer here is

$$\nabla\left(\rho(x,y)\nabla T(x,y;t)\right) = \frac{\partial T(x,y;t)}{\partial t} \quad in \quad \Omega, \tag{7.11}$$

$$T(x,y;0) = T_0 \quad in \quad \Omega, \tag{7.12}$$

$$T(x,y;t) = T_\Gamma \quad on \quad \Gamma, \tag{7.13}$$

for $t \in [0, t_f]$, and where $\rho(x,y)$ is the thermal conductivity, T_0 is the initial temperature and T_Γ is the boundary temperature.

On the other hand, experimental data is obtained from measurements of the temperature for different time steps and at different points on the domain

$$\begin{array}{cccccc} t_1 & \hat{T}_{11} & \hat{T}_{12} & \dots & \hat{T}_{1Q} \\ t_2 & \hat{T}_{21} & \hat{T}_{22} & \dots & \hat{T}_{2Q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_P & \hat{T}_{P1} & \hat{T}_{P2} & \dots & \hat{T}_{PQ} \end{array}$$

where P and Q are the number of time steps and points considered, respectively.

Figure 7.7 is a graphical statement of this situation.

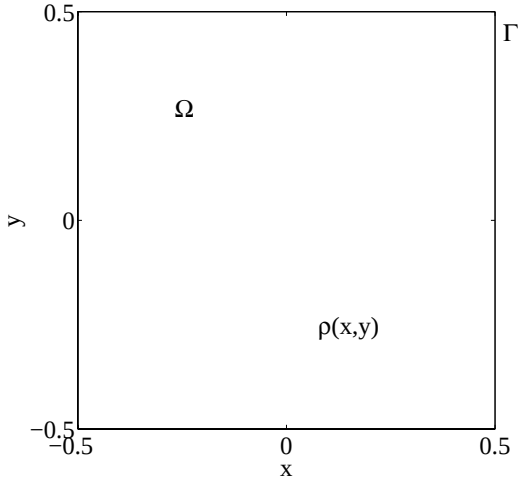


Figure 7.7: The thermal conductivity estimation problem statement.

The goal here is to estimate the thermal conductivity $\rho(x,y)$, so that the heat equation model matches the experimental data. In this way, the objective functional for this problem can be the

mean squared error between the computed temperature for a given thermal conductivity and the measured temperature,

$$E[\rho(x,y)] = \frac{1}{PQ}\sum_{i=1}^{P}\left(\sum_{j=1}^{Q}\left(T(x_j,y_j;t_i) - \hat{T}_{ij}\right)^2\right). \tag{7.14}$$

For this example we take the problem domain to be the square $\Omega = \{(x,y) : |x| \leq 0.5, |y| \leq 0.5\}$ with boundary $\Gamma = \{(x,y) : |x| = 0.5, |y| = 0.5\}$. That domain is discretized with a triangular mesh composed of 888 elements and 485 nodes, and the time is also discretized in 11 time steps. On the other hand, artificial temperature data is generated with the following expression for the thermal conductivity, which will be considered to be the analytical solution to this problem,

$$\rho^*(x,y) = x^2 + y^2. \tag{7.15}$$

Selection of function space

The thermal conductivity $\rho(x,y)$ in Ω is represented by a multilayer perceptron with a sigmoid hidden layer and a linear output layer. The neural network must have two inputs and one output neuron. We guess a good number of neurons in the hidden layer to be three. This neural network is denoted as a $2:3:1$ multilayer perceptron.

It spans a family V of parameterized functions $\rho(x,y;\underline{\alpha})$ of dimension $s = 13$, which is the number of parameters in the neural network. Figure 7.8 is a graphical representation of this network architecture.

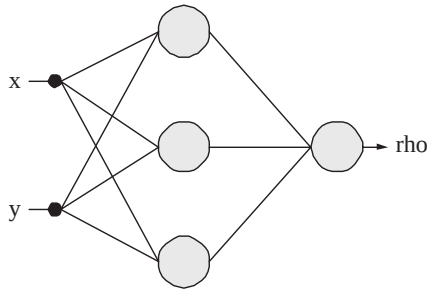


Figure 7.8: Network architecture for the thermal conductivity estimation problem.

Such a multilayer perceptron spans a family V of parameterized functions $\rho(x,y;\underline{\alpha})$ of dimension $s = 19$, which is the number of parameters in the network. Elements V are of the form

$$\begin{aligned} \rho: \quad \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x,y) &\mapsto \rho(x,y;\underline{\alpha}), \end{aligned}$$

where

$$\rho(x,y;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \cdot \tanh\left(b_j^{(1)} + w_{j1}^{(1)}x + w_{j2}^{(1)}y\right). \tag{7.16}$$

On the other hand, the biases and synaptic weight are initialized at random, which means that a random initial guess is used for the thermal conductivity.

Formulation of variational problem

Following Equation 7.14 the variational statement for this properties estimation inverse problem is to find a function $\rho^*(x, y; \underline{\alpha}^*)$ for which the mean squared error functional

$$E[\rho(x, y; \underline{\alpha})] = \frac{1}{PQ} \sum_{i=1}^{P} \left(\sum_{j=1}^{Q} \left(T(x_j, y_j; t_i) - \hat{T}_{ij} \right)^2 \right), \tag{7.17}$$

defined on V takes on a minimum value

Evaluation of the objective functional (7.17) requires a numerical method for integration of partial differential equations. Here we choose the Finite Element Method [102]. For this problem we use a triangular mesh with 888 elements and 485 nodes.

Last, evaluation of the gradient, $\nabla f(\underline{\alpha})$, is carried out by means of numerical differentiation. In particular, the central differences method is used with an ϵ value of 10^{-6} [15] .

Solution of reduced function optimization problem

A suitable training algorithm for this problem is a quasi-Newton method with BFGS train direction and Brent optimal train rate methods for training [83]. Indeed, this training algorithm has demonstrated fully convergence to the global optimum in this problem.

In this example, we set the training algorithm to stop after when the training rate is zero for a gradient descent train direction. During the training process, which lasts for 168 epochs, the error decreases from 1.117 to $2.868 \cdot 10^{-5}$, and the gradient norm from 0.466 to $1.106 \cdot 10^{-6}$. Figures 7.9 and 7.10 show, with a logarithmic scale for the y-axis, the evaluation and gradient norm histories of the neural network training process.

Table 7.2 shows the training results for this problem. Here N denotes the number of training epochs, M the number of objective functional evaluations, CPU the CPU time in seconds for a laptop AMD 3000, $\|\alpha\|$ the final parameters norm, $f(\underline{\alpha}^*)$ the final objective function evaluation, and $\|\nabla f(\underline{\alpha}^*)\|$ its gradient norm.

N	=	168
M	=	10476
CPU	=	16253
$\|\underline{\alpha}^*\|$	=	21.606
$f(\underline{\alpha}^*)$	=	$2.868 \cdot 10^{-5}$
$\|\nabla f(\underline{\alpha}^*)\|$	=	$1.106 \cdot 10^{-3}$

Table 7.2: Training results for the thermal conductivity estimation problem.

The solution here is good, since the mean squared error between the output from the model and the experimental data is of order 10^{-6} and the estimated thermal conductivity matches very well the actual thermal conductivity given by Equation (7.15). Figure 7.11 shows the thermal conductivity estimated by the neural network for this problem.

Finally, the analytical expression of the function represented by the trained multilayer perceptron is as follows,

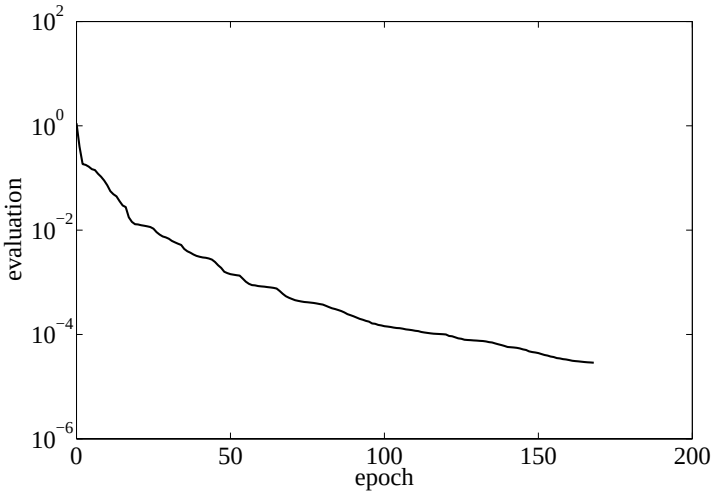


Figure 7.9: Evaluation history for the thermal conductivity estimation problem.

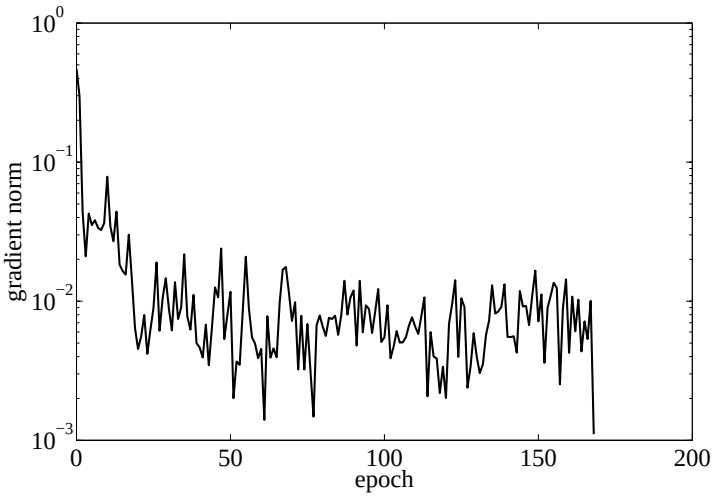


Figure 7.10: Gradient norm history for the thermal conductivity estimation problem.

$$\begin{aligned}\rho^*(x,y;\underline{\alpha}^*) &= 51.6455 \\ &+ 45.7732\tanh\left(-0.551456 - 0.143683x - 0.0824186y\right) \\ &- 30.2139\tanh\left(0.646978 - 0.238919x + 0.0769619y\right) \\ &- 19.8059\tanh\left(0.661936 - 0.00304418x - 0.333647y\right). \end{aligned} \tag{7.18}$$

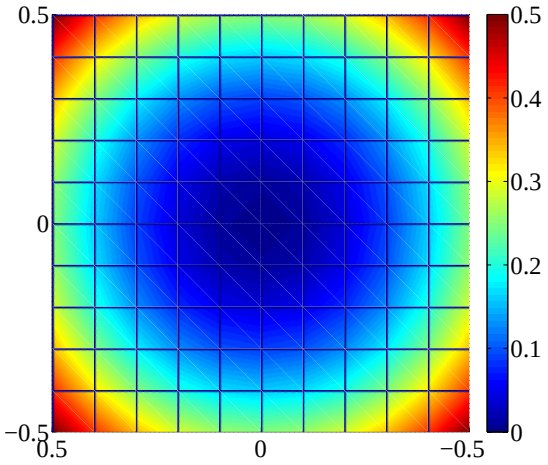


Figure 7.11: Neural network results to the thermal conductivity estimation problem.

7.3 Microstructural modeling of aluminium alloys

This work presents a neural networks approach for modeling the dissolution rate of hardening precipitates in aluminium alloys using inverse analysis. The effective activation energy is also in unison determined as that providing the best results.

As way of illustration, a class of multilayer perceptron extended with independent parameters is trained to find the optimal dissolution curve of hardening precipitates and the effective activation energy of aluminium alloys 2014-T6 and 7449-T79. The results from this Section are published in [65].

7.3.1 Introduction

In heat-treatable aluminium alloys, the dominant microstructural effects are due to changes in the precipitation state of the material, which provide the principal strengthening mechanism [77]. Modeling the dissolution rate of hardening precipitates is therefore of great importance for predicting the hardness distribution after reheating and subsequent natural ageing of the base metal. Subsequent applications of that dissolution model involve a wide range of operational conditions, such as multi-step isothermal heat treatments and continuous thermal cycles [77].

In this regard, various semi-empirical approaches to model the dissolution rate of hardening precipitates in aluminium alloys have proposed in the literature.

The dissolution model of Myhr and Grong [77] is composed from elements derived from thermodynamics, kinetic theory and simple dislocation mechanics, and provides the most important ideas from which this work is developed. It contains a single independent variable, the time, and a single state variable, the volumetric fraction of hardening precipitates. The Myhr and Grong dissolution model provides a basis for obtaining quantitative information of the reaction kinetics through simple hardness measurements, rather than through microscopic observations of the precipitates fraction.

However, Shercliff et al. [92] showed that the dissolution model of Myhr and Grong overestimates the amount of relative volumetric fraction of precipitates at the later stages of this process.

They also proposed a new form for the curve of dissolution.

In this work a novel approach using neural networks is proposed for the identification of microstructural parameters for precipitates dissolution in precipitation hardenable aluminium alloys. The microstructural modeling of aluminium alloys is formulated as a variational problem including independent parameters. This makes it possible to be approached with an extended class of multilayer perceptron to construct the model of different aluminium alloys [68].

The novel methodology is here applied to two different aluminium alloys, 2014-T6 and 7449-T79. The experimental hardness data needed to construct the model is taken from both the literature and through independent measurements.

Section 7.3.2 deals with the physical basis of the dissolution model, which are due to Myhr and Grong. The modeling problem is then formulated as a variational problem with associated independent parameters.

In Section 7.3.3 describes the two materials studied here, aluminium 2014-T6 and aluminium 7449-T79. The experimental data actually employed in the construction of the model is also depicted.

Section 7.3.4 describes the application of the model for the two case studies with a neural network. The dissolution rate of hardening precipitates in these aluminium alloys is modeled, and the effective activation energy is also in unison determined as that providing the best results.

Both the source code and the experimental data used in this work can be found in the open source neural networks C++ library Flood [61].

7.3.2 Dissolution models for aluminium alloys

First, and assuming that the nucleation of precipitates is negligible compared to the dissolution of precipitates, the following linear relationship between the Vickers hardness and the volumetric fraction of precipitates can be established [78]

$$\frac{f}{f_0}(HV) = \frac{HV - \min(HV)}{\max(HV) - \min(HV)}, \tag{7.19}$$

where HV is the actual hardness, $\max(HV)$ is the initial hardness in the hardened state, and $\min(HV)$ is the final hardness after complete dissolution of the precipitates, which can be obtained directly from a Vickers hardness test. Equation (7.19) is extremely useful since the hardness is much easier to measure than the relative volume fraction of precipitates.

On the other hand, as it has been said above, this work is based on the model developed by Myrh and Grong [77] for cylindrical precipitates. This describes the kinetic of transformation of hardening precipitates in aluminium alloys at constant temperature by the following expression,

$$\frac{f}{f_0}(t) = 1 - \sqrt{\frac{t}{t^*(T)}}, \tag{7.20}$$

where f/f_0 is the ratio of the current volume fraction of hardening precipitates f to the initial volume fraction of hardening precipitates f_0, t is the time and $t^*(T)$ is the time for complete dissolution at temperature T. The full dissolution time $t^*(T)$ can be calculated as [77]

$$t^*(T) = t_R \exp\left[\frac{Q}{R}\left(\frac{1}{T} - \frac{1}{T_R}\right)\right], \tag{7.21}$$

where $R = 8.314 \cdot 10^{-3} kJ/molK$ is the universal gas constant, t_R is the reference time for complete dissolution at the reference temperature T_R selected for calibration, and Q is the effective activation energy for precipitates dissolution.

It is easy to see that taking $\log(1 - f/f_0)$ as a function of $\log(t/t^*)$ in Equation (7.20), the resulting plot results in a straight line of slope 0.5.

However, as it was shown by Shercliff et al. [92], the dissolution model of Myrh and Grong overestimates the dissolution rate at the later stages. Shercliff et al. proposed a new model of the form

$$\log(1 - \frac{f}{f_0}) = g(\log(\frac{t}{t^*})). \tag{7.22}$$

where g is a function whose form is given by a "look-up table". Here the early stages of dissolution follow a straight line of gradient 0.5, but this slope steadily decreases at the later stages.

The results from this work will show that although Equation (7.22) seems to give a good estimation at medium and later stages of dissolution, it underestimates the dissolution rate at the lower stages.

The formulation of Shercliff et al. also presents some other drawbacks. First, the modeling process occurs in two steps, estimation of the effective activation energy and generation of the dissolution model. This results in poor accuracy. Second, a "look-up table" is an unsuitable result here. Third, the logarithmic representation in the y-axis produces spurious effects. Indeed, separates the data at the lower stages of dissolution and joins it at the later stages.

The representation of the dissolution model proposed in this work is logarithmic in the x-axis but not in the y-axis. In particular it is of the form

$$1 - \frac{f}{f_0} = g(\log(\frac{t}{t^*})). \tag{7.23}$$

where the function g is to be found from a suitable function space.

The representation for the dissolution model of Equation (7.23) is chosen instead that of Equation (7.22) for a number of reasons: First, it makes the y-axis varying from 0 to 1 and second, it gives all dissolution values the same importance.

The dissolution modeling process is to estimate an activation energy Q providing minimum dispersion for the experimental data while a function $g(\cdot)$ providing minimum error. Mathematically, this can be formulated as a variational problem including independent parameters.

7.3.3 Experimental data

Experimental tests have been performed in order to get the isothermal time evolution of Vickers hardness at different temperatures and for various aluminium alloys. In particular, two materials have been used for the isothermal heat treatments, 2014-T6 and 7449-T79 aluminium alloys.

The Vickers hardness data for aluminium alloy 2014-T6 is taken from [92], while that for aluminium alloy 7449-T79 is obtained from an independent test performed within the DEEPWELD Specific Targeted Research Project (STREP) co-funded by the 6th Framework Programme of the European Community (AST4-CT-2005-516134).

For both types of aluminium, each sample was immersed in a salt bath at different temperatures ranging from $200°C$ to $400°C$ during periods ranging from 2 seconds to about 1 day. The samples were then water quenched. Immediately after, the surface was polished and about five

Vickers hardness measurements were performed at room temperature on each sample. The hardness measurements were repeated one week later to observe ageing. The total number of samples for aluminium 2014-T6 is 45, and the total number of samples aluminium 7449-T79 is 70.

Figures 7.12 and 7.13 depict these Vickers hardness test for aluminium alloys 2014-T6 and AA7449-T6, respectively. Note that, in both figures the Vickers hardness decreases with the time, due to dissolution of hardness precipitates.

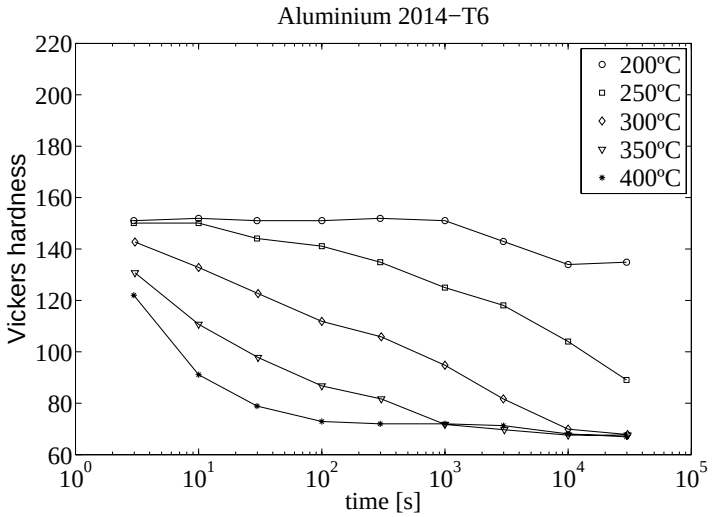


Figure 7.12: Vickers hardness test for aluminium alloy 2014-T6.

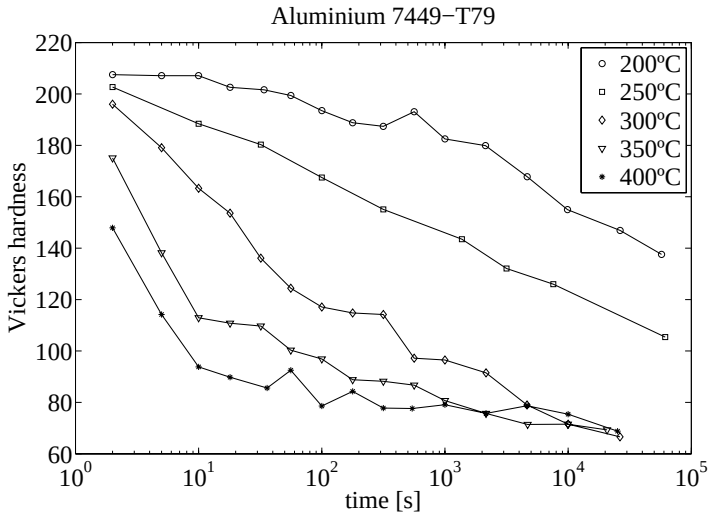


Figure 7.13: Vickers hardness test for aluminium alloy 7449-T79.

Table 7.3 shows the parameters used here for both aluminium alloys needed to transform

from Vickers hardness to volumetric fraction of hardening precipitates in Equation (7.19). Here $\min(HV)$ is the minimum Vickers hardness, $\max(HV)$ is the maximum Vickers hardness, $t_R[s]$, is the reference time in seconds and $T_R[K]$ is the reference temperature in Kelvin.

Aluminium	$\min(HV)$	$\max(HV)$	$t_R[s]$	$T_R[K]$
2014-T6	65	155	10000	573.16
7449-T79	63	207.5	16	623.16

Table 7.3: Parameters for aluminium alloys 2014-T6 and 7449-T79.

7.3.4 Numerical results

A variational formulation for neural networks is in this section applied to find the optimal dissolution model and effective activation energy of aluminium alloys 2014-T6 and 7449-T79.

Selection of function space

The first step is to choose a function space to represent the dissolution model $(1-f/f_0)(\log(t/t^*))$. Here a multilayer perceptron with a sigmoid hidden layer and a linear output layer is used. This neural network is very useful in inverse problems, since it is a class of universal approximator [49].

The number of inputs and output neurons are constrained by the problem. Here the multilayer perceptron must have one input, $\log(t/t^*)$, and one output neuron, $1-f/f_0$. However, the number of hidden neurons is a design variable in the problem.

In order to draw a correct network architecture for the dissolution model, different sizes for the hidden layer are compared and that providing best validation performance is chosen. In particular, the Vickers hardness tests are divided into training and validation subsets. A multilayer perceptron with 1, 2 and 3 hidden neurons are then trained with the training data and their performance against the validation data are compared. It is found that the optimal network architecture for the two case studied considered here is that with 1 hidden neuron.

Overall, the network architecture used to represent the dissolution model for aluminium 2014-T6 and aluminium 7449-T79 has one input, one sigmoid hidden neuron and one linear output neuron. This is the simplest multilayer perceptron possible. Figure 7.14 is a graphical representation of this network architecture.

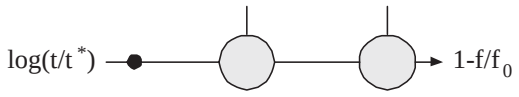


Figure 7.14: Network architecture for aluminium alloys 2014-T6 and 7449-T79.

On the other hand, and in order to obtain a correct representation for the solution, information about the effective activation energy is required. Thus, an independent parameter Q must be associated to the multilayer perceptron. Note that a standard software tool which implements the multilayer perceptron will not be capable of dealing with this problem, since it needs to include independent parameters.

This extended class of multilayer perceptron spans a family V of functions $(1-f/f_0)(\log(t/t^*;\underline{\alpha},Q)$ of dimension $d=4+1$, where 4 is the number of biases and synaptic weights and 1 is the number of independent parameters. Elements of V are of the form

$$\begin{aligned} 1-\frac{f}{f_0}: \quad \mathbb{R} &\rightarrow \mathbb{R} \\ \log(\frac{t}{t^*}) &\mapsto (1-\frac{f}{f_0})(\log(\frac{t}{t^*});\underline{\alpha},Q). \end{aligned}$$

Here the biases and synaptic weights, $\underline{\alpha}$, are initialized at random in the interval $[-1,1]$, and the activation energy, Q, is also initialized at random in the interval $[100,200]$. In this way a random model with a random value for the activation energy is used as an initial guess.

Formulation of variational problem

The second step is to select an appropriate objective functional, in order to formulate the variational problem. A Minkowski error could be used here to perform robust modeling. Also, a regularization term could be included to prevent overfitting. However, the Vickers hardness tests for both alloys do not show the presence of significant outliers. Also, the network architecture chosen is that providing the minimum validation error. So instead, the simpler mean squared error between the dissolution model and the volumetric fraction data is used. The mathematical expression of this objective functional is

$$F\left[(1-\frac{f}{f_0})(\log(\frac{t}{t^*});\underline{\alpha},Q)\right]=\frac{1}{S}\sum_{s=1}^{S}\left((1-\frac{f}{f_0})(\log(\frac{t}{t^*});\underline{\alpha},Q)^{(s)}-(1-\frac{f}{f_0})^{(s)}\right)^2 \tag{7.24}$$

where S is the number of samples available, and $(1-f/f_0)^{(s)}$ is calculated by means of Equation 7.19.

Note that, in Equation (7.24), the pairs of values $(\log(t/t^*),1-f/f_0)$ are not fixed, since they depend on the particular value used for the effective activation energy Q. In this way, the optimal value for this independent parameter will provide the minimum dispersion of the data. As a consequence, a standard software tool implementing the mean squared error will not be applicable to this problem. Instead, programming the objective functional for this particular application is required.

The variational problem can then be formulated so as to find a function $(1-f/f_0)^*(\log(t/t^*);\underline{\alpha}^*,Q^*)$ and an independent parameter Q^* for which the objective functional in Equation (7.24) takes on a minimum value.

On the other hand, evaluation of the objective function gradient vector, $\nabla f(\underline{\alpha},Q)$, is performed via numerical differentiation. In particular, the central differences method is used with an epsilon value of 10^{-6}.

Solution of reduced function optimization problem

The third step is to choose a suitable training algorithm to solve the reduced function optimization problem. To check for presence of local minima a quasi-Newton method with BFGS train direction and Brent optimal train rate methods has been used from different initial guesses. The results provided are in all cases the same, and no local minima seem to appear for these particular case studies.

For both aluminium alloys, the training algorithm is set to stop when the quasi-Newton method cannot perform any better. At this point the algorithm gives zero optimal train rate in the Brent method for some gradient descent train direction. Figures 7.15 and 7.16 depict the training history for aluminium alloys 2014-T6 and 7449-T79, respectively. Note that a base 10 logarithmic scale is used for the y-axis in both plots. They depict the mean squared error value and its gradient norm as a function of the training epoch. In both figures the evaluation decreases very fast during the first few epochs to reach a stationary value. Also, the gradient norm approaches a very small value. These two signs indicate that the algorithm has fully converged to a minimum.

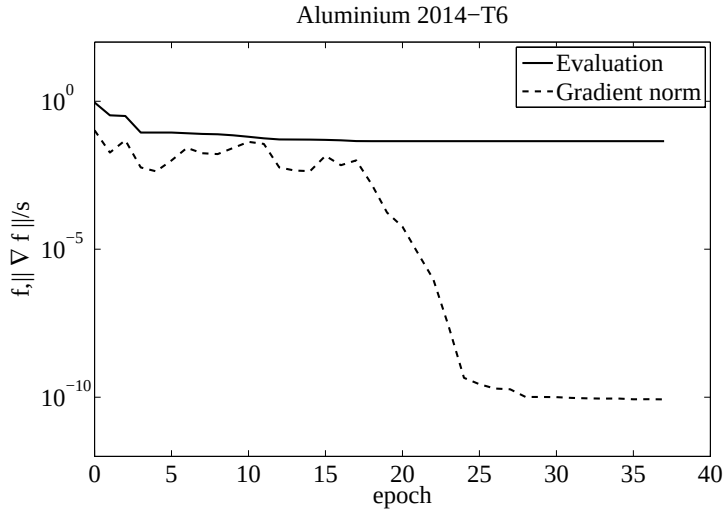


Figure 7.15: Training history for aluminium alloy 2014-T6.

Table 7.4 shows the training results for these case studies. Here N denotes the number of epochs, M the number of evaluations, CPU the CPU time in a laptop AMD 3000, $f(\underline{\alpha}^*, Q^*)$ the final objective function value, $\nabla f(\underline{\alpha}^*, Q^*)$ its gradient norm and Q^* the optimum effective activation energy.

Aluminium	N	M	CPU	$f(\underline{\alpha}^*, Q^*)$	$\nabla f(\underline{\alpha}^*, Q^*)$	Q^*
2014-T6	38	1614	$1\,s$	$4.465 \cdot 10^{-2}$	$8.453 \cdot 10^{-11}$	151.636
7449-T79	27	1282	$1\,s$	$5.398 \cdot 10^{-2}$	$9.854 \cdot 10^{-10}$	145.458

Table 7.4: Training results for aluminium alloys 2014-T6 and 7449-T79.

The mathematical expression of the dissolution model provided by the neural network for aluminium 2014-T6 is given by

$$(1 - \frac{f}{f_0})(\log(\frac{t}{t^*}); \underline{\alpha}^*, Q^*) = \frac{1}{2}\Bigg\{0.0324727 + 0.959812 \tanh\left[1.0037 + 3.67865\left(2\frac{\log(t/t^*) + 6}{12} - 1\right)\right] + 1\Bigg\}. \tag{7.25}$$

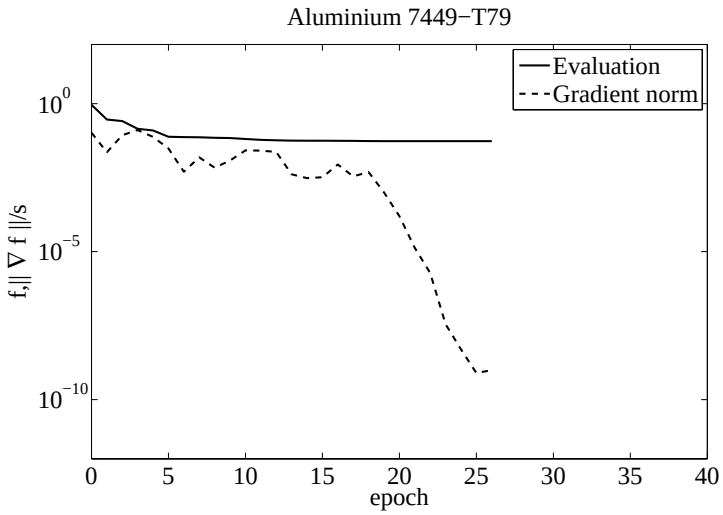


Figure 7.16: Training history for aluminium alloy 7449-T79.

with $Q^* = 151.636$. The mathematical expression for aluminium 7449-T79 is

$$(1 - \frac{f}{f_0})(\log(\frac{t}{t^*}); \underline{\alpha}^*, Q^*) = \frac{1}{2}\Bigg\{ - 0.0159086 + 0.974911 \tanh\left[0.259412 + 3.15452\left(2\frac{\log(t/t^*) + 6}{12} - 1\right)\right] + 1\Bigg\}. \tag{7.26}$$

with $Q^* = 145.458$.

Figures 7.17 and 7.18 show the dissolution models for aluminium 2014-T6 and aluminium 7449-T79, respectively. Comparing the shape of these curves to those reported in the literature there are similarities and differences. For medium and high stages of dissolution they are very similar tho those reported by Shercliff et al. [92]. However, these results show that the dissolution model Shercliff et al. underestimates the dissolution rate at the lower stages, while the neural networks method does not.

7.3.5 Conclusions

Neural networks were successfully used to calibrate a simple microstructure model based on dissolution of the hardening precipitates.

Future work includes applying of the dissolution model for process simulation in the metallurgical industry, such as friction stir welding.

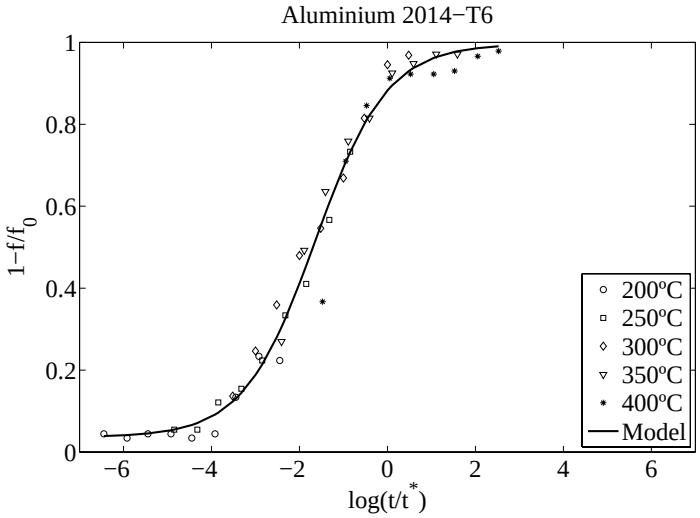


Figure 7.17: Dissolution model for aluminium alloy 2014-T6.

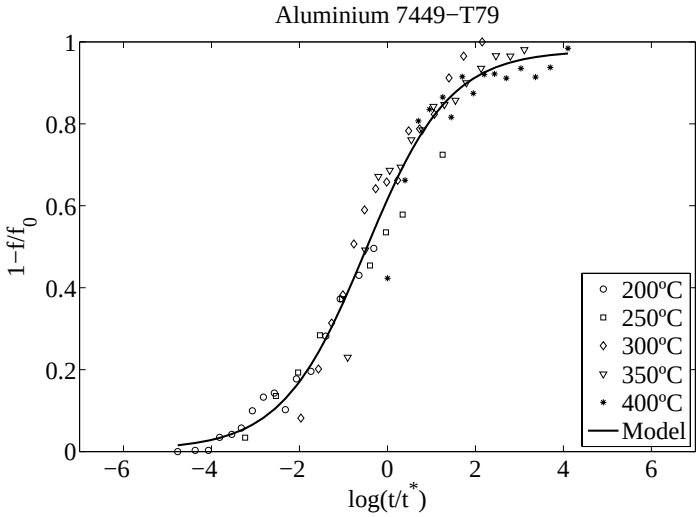


Figure 7.18: Dissolution model for aluminium alloy 7449-T79.

This page is intentionally left blank.

Chapter 8

Optimal shape design

This section includes a short survey to optimal shape design, a theory which is developed from the abstract field of calculus of variations. In this way, optimal shape design is here stated as a new learning task for the multilayer perceptron.

The application of this approach is validated through an optimal shape design problem whose exact solution is known. More practical implementation issues are studied by using a neural network to design an airfoil with given thickness and providing maximum aerodynamic efficiency for transonic flight conditions.

8.1 Mathematical formulation

Optimal shape design is a very interesting field both mathematically and for industrial applications. The goal here is to computerize the design process and therefore shorten the time it takes to design or improve some existing design. In an optimal shape design process one wishes to optimize a criteria involving the solution of some mathematical model with respect to its domain of definition [74]. The detailed study of this subject is at the interface of variational calculus and numerical analysis.

In order to properly define an optimal shape design problem the following concepts are needed:

(i) A mathematical model of the system.

(ii) A statement of the physical constraints.

(iii) A specification of the performance criterion.

Mathematical model

The mathematical model is a well-formed formula which involves the physical form of the device to be optimized. Let define $\mathbf{y}(\mathbf{x})$ the shape variables and $\mathbf{u}(\mathbf{x})$ the state variables. The mathematical model or state equation can then be written as

$$\mathcal{L}(\mathbf{y}(\mathbf{x}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}, \tag{8.1}$$

where $\mathcal{L}$ is some algebraic or differential operator and $\mathbf{f}$ some forcing term.

Physical constraints

An optimal shape design problem might also be specified by a set of constraints on the shape and the state variables of the device.

Two important types of shape constraints are boundary conditions and lower and upper bounds. If some outputs are specified for given inputs, then the problem is said to include boundary conditions. On the other hand, if some shape variables are restricted to fall in some interval, then the problem is said to have lower and upper bounds.

State constraints are conditions that the solution to the problem must satisfy. This type of constraints vary according to the problem at hand.

In this way, a design which satisfies all the shape and state constraints is called an admissible shape.

Definition 14 (Admissible shape)**.** A shape $\mathbf{y}(\mathbf{x})$ which satisfies all the constraints is called an admissible shape. The set of admissible shapes is denoted Y, and the notation $\mathbf{y}(\mathbf{x}) \in Y$ means that $\mathbf{y}(\mathbf{x})$ is admissible.

Similarly, a state which satisfies the constraints is called an admissible state.

Definition 15 (Admissible state)**.** A state $\mathbf{u}(\mathbf{x})$ which satisfies the state variable constraints is called an admissible state. The set of admissible states will be denoted by $\mathbf{U}$, and $\mathbf{u}(\mathbf{x}) \in \mathbf{U}$ means that the state $\mathbf{u}(\mathbf{x})$ is admissible.

Performance criterion

The performance criterion expresses how well a given design does the activity for which it has been built. In optimal shape design the performance criterion is a functional of the form

$$\begin{array}{rccc} F: & Y & \to & \mathbb{R} \\ & \mathbf{y}(\mathbf{x}) & \mapsto & F[\mathbf{y}(\mathbf{x})]. \end{array}$$

Optimal shape design problems solved in practice are, as a rule, multi-criterion problems. This property is typical when optimizing the device as a whole, considering, for example, weight, operational reliability, costs, etc. It would be desirable to create a device that has extreme values for each of these properties. However, by virtue of contradictory of separate criteria, it is impossible to create devices for which each of them equals its extreme value.

To sum up, the optimal shape design problem can be formulated as

Problem 11 (Optimal shape design problem)**.** Let Y and U be the function spaces of admissible shapes and states, respectively. Find an admissible shape $\mathbf{y}^*(\mathbf{x}) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to be in an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$ and for which the performance criterion

$$F[\mathbf{y}(\mathbf{x})]$$

takes on a minimum or maximum value. The function $\mathbf{y}^*(\mathbf{x})$ is called an optimal shape and the function $\mathbf{u}^*(\mathbf{x})$ an optimal state.

Solution methods

In general, there are no automatic solutions to optimal shape design problems. Therefore, the use of direct methods usually becomes necessary.

A variational formulation for neural networks provides a direct method for the solution of variational problems. Therefore optimal shape design problems can be approached with this numerical technique. From the nomenclature introduced in this section, the optimal shape design problem for the multilayer perceptron can be written as follows:

Problem 12 (Optimal shape design problem for the multilayer perceptron)**.** Let Y and U be the function spaces of admissible shapes and states, respectively. Let also V be the space consisting of all shapes $\mathbf{y}(\mathbf{x};\underline{\alpha})$ that a given multilayer perceptron can define, with dimension s. Find an admissible shape $\mathbf{y}^*(\mathbf{x};\underline{\alpha}^*) \in Y$ which causes the system

$$\mathcal{L}(\mathbf{y}(\mathbf{x};\underline{\alpha}), \mathbf{u}(\mathbf{x}), \mathbf{x}) = \mathbf{f}$$

to be in an admissible state $\mathbf{u}^*(\mathbf{x}) \in U$, and for which the performance criterion

$$F[\mathbf{y}(\mathbf{x};\underline{\alpha})],$$

defined on V, takes on a minimum or a maximum value.

8.2 Validation examples

In this section the application of optimal shape design problems for the multilayer perceptron is validated by means of an unconstrained and a constrained problem. These case studies can be solved analytically, which enables to compare the neural network results against the exact ones.

8.2.1 The minimum drag problem

The minimum drag problem for the multilayer perceptron is an optimal shape design problem with one input and one output variables, besides two boundary conditions. It is defined by an unconstrained objective functional requiring the integration of a function. This problem is included with the Flood library [61], and it has been published in [64] and [63].

Problem statement

Consider the design of a body of revolution with given length l and diameter d providing minimum drag at zero angle of attack and for neglected friction effects, see Figure 8.1.

The drag of such a body $y(x)$ can be expressed as

$$D[y(x)] = 2\pi q \int_0^l y(x)[C_p y'(x)]dx, \tag{8.2}$$

where q is the free-stream dynamic pressure and C_p the pressure coefficient [17]. For a slender body $y'(x) \ll 1$, the pressure coefficient can be approximated by the Newtonian flow relation

$$C_p = 2\left[y'(x)\right]^2, \tag{8.3}$$

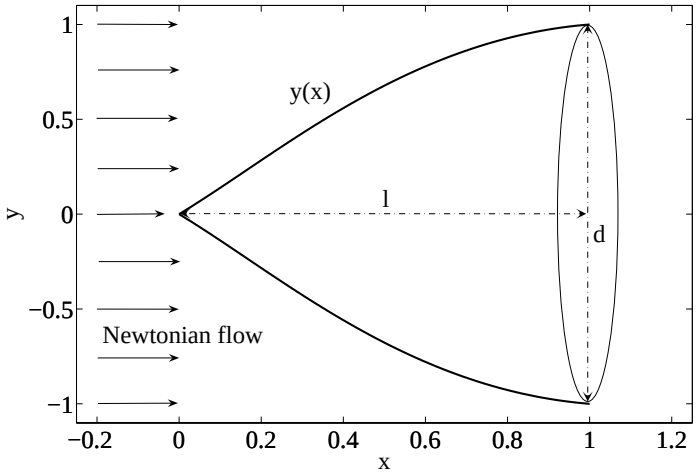


Figure 8.1: The minimum drag problem statement.

which is valid provided that the inequality $y'(x) \geq 0$ is satisfied.

From Equations (8.2) and (8.3) we obtain the following approximation for the drag,

$$D[y(x)] = 4\pi q \int_0^l y(x)[y'(x)]^3 dx. \tag{8.4}$$

It is convenient to introduce the following dimensionless variables associated with the axial coordinate and the radial coordinate

$$\xi = \frac{x}{l}, \tag{8.5}$$

$$\eta = \frac{2y}{d}. \tag{8.6}$$

In that way, both ξ and η vary from 0 to 1.

Also, a dimensionless coefficient associated with the drag can be defined as

$$C_D[\eta(\xi)] = \tau^2 \int_0^1 \eta(\xi)[\eta'(\xi)]^3 d\xi, \tag{8.7}$$

where $\tau = d/l$ is the slenderness of the body.

The analytical solution to the minimum drag problem formulated in this section is given by

$$\eta^*(\xi) = \xi^{3/4}, \tag{8.8}$$

which provides a minimum value for the drag coefficient $C_D/\tau^2 = 0.4220$.

Selection of function space

The body of revolution $\eta(\xi)$, for $\xi \in [0,1]$, will be represented by a multilayer perceptron with a sigmoid hidden layer and a linear output layer. This axisymmetric structure is to be written in cartesian coordinates, so the neural network must have one input and one output neuron. On the other hand, an appropriate number of hidden neurons is believed to be three for this particular application. This network architecture is depicted in Figure 8.2.

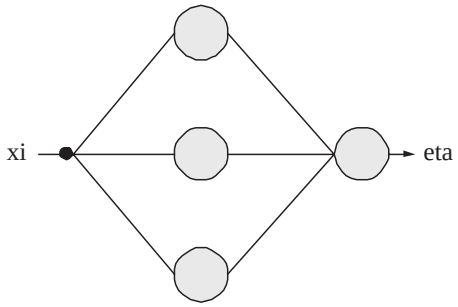


Figure 8.2: Network architecture for the minimum drag problem.

Such a multilayer perceptron spans a family V of parameterized functions $\eta(\xi;\underline{\alpha})$ of dimension $s = 10$, which is the number of parameters in the network. The elements of this function space are of the form

$$\begin{array}{rccc} \eta: & \mathbb{R} & \rightarrow & \mathbb{R} \\ & \xi & \mapsto & \eta(\xi;\underline{\alpha}), \end{array}$$

where

$$\eta(\xi;\underline{\alpha}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \cdot \tanh\left(b_j^{(1)} + w_{j1}^{(1)}\xi\right). \tag{8.9}$$

The outputs from the neural network in Figure 8.2 must hold the boundary conditions $\eta(0) = 0$ and $\eta(1) = 1$. A suitable set of particular and homogeneous solution terms here is

$$\varphi_0(\xi) = \xi, \tag{8.10}$$
$$\varphi_1(\xi) = \xi(\xi - 1), \tag{8.11}$$

respectively. This gives

$$\eta(\xi;\underline{\alpha}) = \xi + \xi(\xi - 1)\eta(\xi;\underline{\alpha}). \tag{8.12}$$

Also, the functions $\eta(\xi)$ are constrained to lie in the interval $[0,1]$. To deal with such constraints the neural network outputs are bounded in the form

$$\eta(\xi;\underline{\alpha}) = \begin{cases} 0, & \eta(\xi;\underline{\alpha}) < 0. \\ \eta(\xi;\underline{\alpha}), & 0 \leq \eta(\xi;\underline{\alpha}) \leq 1. \\ 1, & \eta(\xi;\underline{\alpha}) > 1. \end{cases} \tag{8.13}$$

The elements of the function space constructed so far indeed satisfy the boundary conditions and the input constraints. Also, they are thought to have a correct complexity for this case study.

Experience has shown that this method does not require a good initial guess for the solution, so the parameters in the neural network are initialized at random. This is a potential advantage over other direct methods, in which a good initial guess for the solution might be needed. Figure 8.3 depicts the starting random shape for this problem.

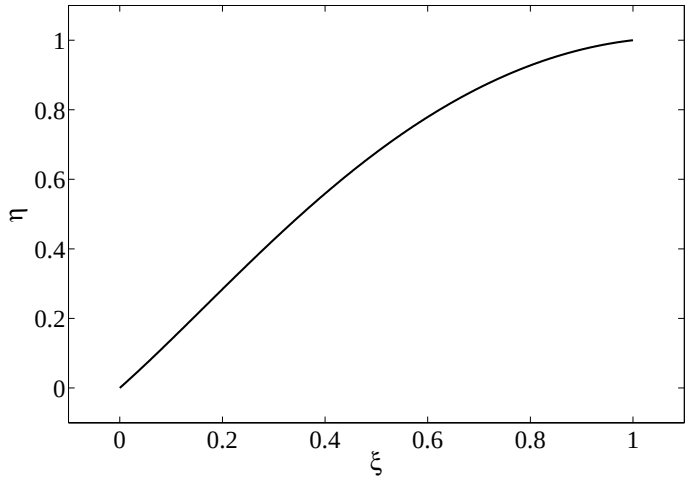


Figure 8.3: Initial guess for the minimum drag problem.

Formulation of variational problem

From Equation (8.7), the variational statement of this problem is to find a function $\eta^*(\xi;\underline{\alpha}^*) \in V$ for which the functional

$$C_D[\eta(\xi;\underline{\alpha})]/\tau^2 = \int_0^1 \eta(\xi;\underline{\alpha})[\eta'(\xi;\underline{\alpha})]^3 d\xi, \tag{8.14}$$

defined on V, takes on a minimum value.

In order to evaluate the objective functional in Equation (8.14) the integration of a function is needed. Here we apply the Runge-Kutta-Fehlberg method [93] with tolerance 10^{-6}.

Solution of reduced function optimization problem

Here we use a quasi-Newton method with BFGS train direction and Brent optimal train rate methods for training [15]. The tolerance in the Brent's method is set to 10^{-6}.

The objective function gradient vector $\nabla f(\underline{\alpha})$ is calculated by means of numerical differentiation. In particular, the symmetrical central differences method is used with $\epsilon = 10^{-6}$ [15].

The evaluation and the gradient norm of the initial guess are 0.56535 and 0.324097, respectively. Trained is performed until the algorithm can not perform any better, that is, when the Brent's method gives zero train rate for a gradient descent train direction. This occurs after 759 epochs, at which the objective function evaluation is 0.422. At this point the gradient norm takes a value of $2.809 \cdot 10^{-4}$. Figures 8.4 and 8.5 illustrate this training process.

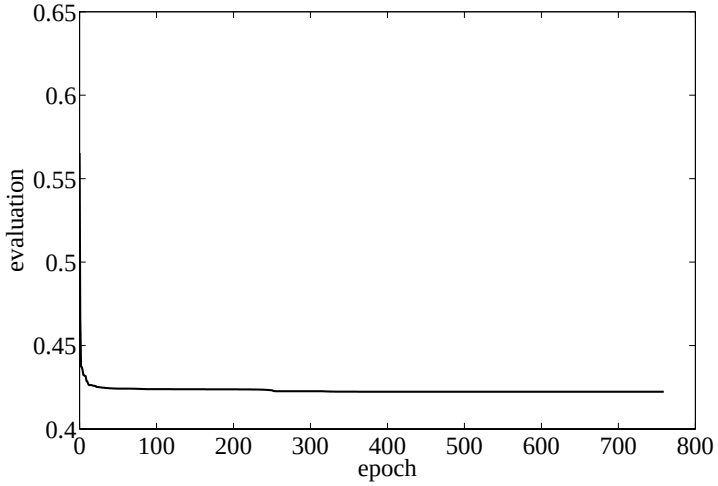


Figure 8.4: Evaluation history for the minimum drag problem.

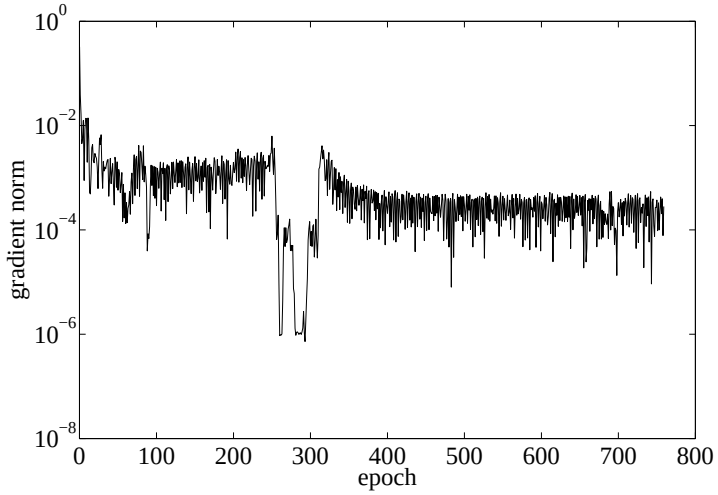


Figure 8.5: Gradient norm history for the minimum drag problem.

Table 8.1 shows the training results for this problem. Here N is the number of training epochs, M the number of objective function evaluations, CPU the CPU time for a laptop AMD 3000, $\|\underline{\alpha}^*\|$ the final parameters norm, $f(\underline{\alpha}^*)$ the final value for the objective function and $\|\nabla f(\underline{\alpha}^*)\|$ the final gradient norm.

N	=	759
M	=	38426
CPU	=	354
$\|\underline{\alpha}^*\|$	=	122.752
$f(\underline{\alpha}^*)$	=	0.422325
$\|\nabla f(\underline{\alpha}^*)\|$	=	$2.80939 \cdot 10^{-4}$

Table 8.1: Training results for the minimum drag problem.

Comparing the drag coefficient provided by that neural network (0.4223) to that by the analytical result (0.4220), these two values are almost the same. In particular, the percentage error made by the numerical method is less than 0.1%.

The optimal shape design by the multilayer perceptron is depicted in Figure 8.6.

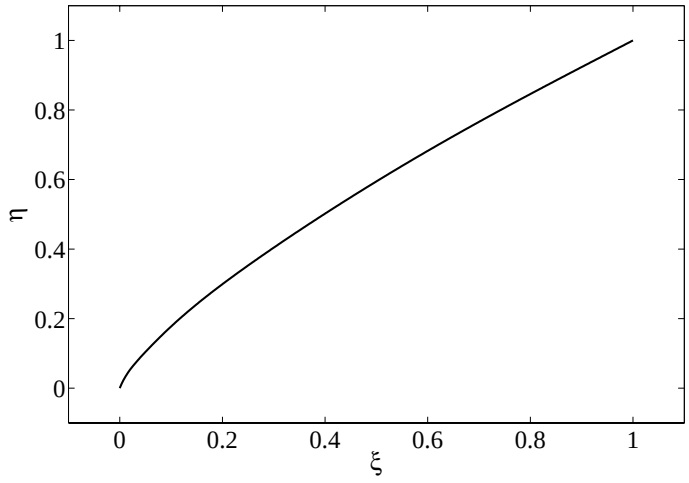


Figure 8.6: Neural network results to the minimum drag problem.

Finally, an explicit expression of the shape of such an axisymmetric body is given by

$$\begin{aligned}\eta^*(\xi;\underline{\alpha}^*) &= \xi + \xi(\xi-1)[-164.639 \\ &- 275.014\tanh(-2.97601 - 27.2435\xi) \\ &- 79.9614\tanh(-2.62125 - 3.7741\xi) \\ &+ 201.922\tanh(-1.78294 + 0.0113036\xi)]. \end{aligned} \tag{8.15}$$

8.3 Optimal airfoil design

A neural networks approach is proposed in this work for the designing of airfoil profiles with maximum aerodynamic efficiency. A practical application is solved by integrating the software tools Flood [61], GiD [25] and PUMI [79] for neural networks, pre and postprocessing and computational fluid dynamics, respectively.

8.3.1 Introduction

The aerodynamic efficiency is the amount of lift generated by a wing or vehicle compared to the drag it creates by moving through the air [43] [44]. It can therefore be defined as the C_l/C_d ratio, where C_l and C_d are the lift and drag coefficients, respectively.

Improvement of the efficiency in one of the major goals in wing design. Indeed, since a given aircraft's needed lift does not change, delivering that lift with lower drag leads directly to better fuel economy, climb performance and glide ratio. In order to obtain high efficiency it is essential to design airfoils using shape optimization techniques [18]. The use of neural networks within a variational formulation provides a direct method for the solution of optimal shape design problems [63].

Neural networks have been widely applied in aeronautics. For instance, an aerodynamic design procedure that incorporates that computational tool is described in [84]. However, the approach here is different since an extended class of neural network is used to define the airfoil shape, and not a response surface for the objective function.

The method is applied to the designing of the shape and the flight conditions of a transonic airfoil. The results demonstrate that the approach is very useful for optimization, since the constructed neural network is able to represent an extensive family of airfoils for any flight condition.

8.3.2 Problem statement

An airfoil can be generated using analytical equations that describe the camber distribution of the airfoil section, $y_c(x)$, as well as the thickness distribution, $y_t(x)$, for $x \in [0, 1]$.

In order to determine the final coordinates for the airfoil upper surface (x_U, y_U) and lower surface (x_L, y_L) the following relationships are used

$$x_U = x - y_t \sin(\theta), \tag{8.16}$$
$$y_U = y_c + y_t \cos(\theta), \tag{8.17}$$
$$x_L = x + y_t \sin(\theta), \tag{8.18}$$
$$y_L = y_c - y_t \cos(\theta), \tag{8.19}$$

where

$$\theta = \arctan\left(\frac{dy_c}{dt}\right). \tag{8.20}$$

A suitable airfoil must hold some conditions for the camber and the thickness distributions, as well as for their derivatives.

The camber distribution must be zero at both borders of the airfoil. This can be written

$$y_c(0) = 0, \tag{8.21}$$
$$y_c(1) = 0. \tag{8.22}$$

On the other hand, the thickness must be zero at the left and right borders, and it must take the semi-maximum thickness value $m/2$ at the position of maximum thickness p,

$$y_t(0) = 0, \tag{8.23}$$
$$y_c(p) = m/2, \tag{8.24}$$
$$y_c(1) = 0. \tag{8.25}$$

The airfoil leading edge must be blunt, with a radius of curvature large enough to prevent boundary layer separation. This is roughly equivalent to the following condition

$$y_t'(0) = \infty. \tag{8.26}$$

At the position of maximum thickness, the derivative of the thickness distribution must be zero

$$y_t'(p) = 0. \tag{8.27}$$

Finally, the trailing edge is sharp, forcing the flow to detach smoothly at the rear end of the section. This can be written

$$y_t'(1) \neq \infty. \tag{8.28}$$

The aim is to determine the shape $(y_c, y_t)(x)$ and the angle of attack, α, of an airfoil under transonic flight conditions, providing maximum aerodynamic efficiency. The maximum thickness p is given but the position of maximum thickness can vary in the interval $m \in [1.0, 0.6]$. The flow is assumed to be Eulerian [79].

The maximum thickness is set here to be $m = 1.5$. In the flow model, the free-stream Mach number is set to $M_\infty = 0.85$.

8.3.3 Numerical results

The optimal shape design problem stated in the preceding subsection is approached in this one using a variational formulation for the multilayer perceptron.

Selection of function space

The network architecture used here is a multilayer perceptron with a sigmoid hidden layer and a linear output layer. It must have one input, x, and two output neurons, y_c and y_t. Three neurons are used in the hidden layer, i.e. a $1:3:2$ multilayer perceptron layer is used here as shown in Figure 8.7.

It is very desirable to include as much a priory information of the problem as possible, in order to span a more suited function space here.

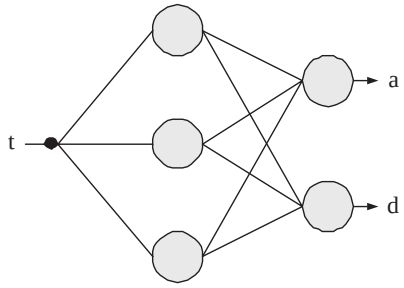


Figure 8.7: Network architecture for the designing of an optimal airfoil.

In this way, two independent parameters are associated to the neural network, the position of the maximum thickness p and the angle of attack α. The total set of free parameters is thus $\underline{\beta} = (\underline{\omega}, p, \alpha)$, where $\underline{\omega}$ is the vector of all the biases and synaptic weights in the neural network.

A few steps must be performed in order to span a proper function space for this problem. These are detailed in the following points.

1. Preprocess input

The inputs to the neural network are first preprocessed with the minimum and maximum method in get input signals lying in the interval $x \in [-1, 1]$. For that purpose we use the values $\min(x) = 0$ and $\max(x) = 1$,

$$x = 2\frac{x - \min(x)}{\max(x) - \min(x)} - 1, \tag{8.29}$$

where $\min(x) = 0$ and $\max(x) = 1$ are the minimum and the maximum values of the input variable, respectively.

2. Obtain output signals

The explicit expression for the output signals by the neural network diagram in Figure 8.7 is of the form

$$y_c(x; \underline{\beta}) = b_1^{(2)} + \sum_{j=1}^{3} w_{1j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} x\right), \tag{8.30}$$

$$y_t(x; \underline{\beta}) = b_2^{(2)} + \sum_{j=1}^{3} w_{2j}^{(2)} \tanh\left(b_j^{(1)} + w_{j1}^{(1)} x\right). \tag{8.31}$$

3. Postprocess output signals

The output signals from the neural network are postprocessed so as to obtain outputs of the form $y_c(x) \in [\min(y_c), \max(y_c)]$ and $y_t(x) \in [\min(y_t), \max(y_t)]$, approximately. This for of post-processing is given by

$$y_c = 0.5\,(y_c + 1)\,(\max(y_c) - \min(y_c)) + \min(y_c), \tag{8.32}$$

$$y_t = 0.5\,(y_t + 1)\,(\max(y_t) - \min(y_t)) + \min(y_t), \tag{8.33}$$

$$\tag{8.34}$$

where $\min(y_c)$, $\max(y_c)$, $\min(y_t)$ and $\max(y_t)$ are an estimation of the minimum and the maximum values of the output variables. A good set for these values depends on the particular and homogeneous terms used to satisfy the boundary conditions. Here we have chosen -0.01, 0.01, -0.001 and 0.001, respectively.

4. Apply boundary conditions

The outputs from the neural network in Figure 8.7 must hold the conditions (8.21), (8.22), (8.23), (8.24), (8.25), (8.26), (8.27) and (8.28). A suitable set of particular and homogeneous solutions for the camber and the thickness, $\varphi_{0c}(x)$, $\varphi_{0t}(x)$, $\varphi_{1c}(x)$ and $\varphi_{1t}(x)$, distributions is thus chosen.

A particular solution for the camber distribution can the mean center line,

$$\varphi_{0c}(x) = 0, \tag{8.35}$$

which indeed satisfies $\varphi_{0c}(0) = 0$ and $\varphi_{0c}(1) = 0$.

The selected particular solution for the thickness distribution is an ellipse with semimajor axis p, semiminor axis $m/2$ and centered at the point $(0, p)$ for $x \in [0, p]$, and a parabola with directrix parallel to the y-axis, vertex (p, m) and passing through $(1, 0)$ for $x \in [p, 1]$.

$$\varphi_{0t}(x) = \begin{cases} \frac{m}{2p}\sqrt{x(2p-x)}, & x \leq p. \\ -\frac{(x^2-2px+2p-1)m}{2(p-1)^2}, & x > p. \end{cases} \tag{8.36}$$

which certainly holds the conditions $\varphi_{0t}(0) = 0$, $\varphi'_{0t}(0) = \infty$, $\varphi_{0t}(p) = m$, $\varphi'_{0t}(p) = 0$, $\varphi_{0t}(1) = 0$ and $\varphi'_{0t}(1) \neq \infty$.

Figure 8.8 shows the upper and lower coordinates for the particular solution, which is considered to be the baseline airfoil in the designing process.

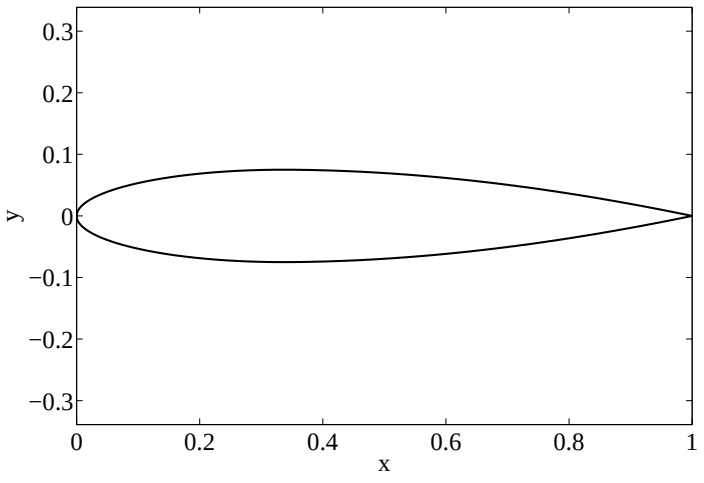


Figure 8.8: Baseline for the designing of an optimal airfoil.

On the other hand, it is easy to find an homogeneous solution for the camber as

$$\varphi_{1c}(x) = \sin(\pi x), \tag{8.37}$$

which satisfies $\varphi_{1c}(0) = 0$ and $\varphi_{1c}(1) = 0$, and varies between 0 and 1.

Similarly, an homogeneous solution for the thickness is chosen to be

$$\varphi_{1t}(x) = \begin{cases} 0.5 - 0.5\sin\left(\frac{\pi}{2} + 2\pi\frac{x}{p}\right), & x \leq p. \\ -\frac{27}{4(p-1)^3}(x-p)^2(x-1), & x > p. \end{cases} \tag{8.38}$$

which holds $\varphi_{1t}(0) = 0$, $\varphi'_{1t}(0) = 0$, $\varphi_{1t}(p) = 0$, $\varphi'_{1t}(p) = 0$, $\varphi_{1t}(1) = 0$ and $\varphi'_{1t}(1) \neq 0$. This function varies between -1 and 1.

The expressions for the camber and thickness distributions then become

$$y_c(x;\underline{\beta}) = \varphi_{0c}(x) + \varphi_{1c}(x)y_c(x;\underline{\beta}), \tag{8.39}$$

$$y_t(x;\underline{\beta}) = \varphi_{0t}(x) + \varphi_{1t}(x)y_t(x;\underline{\beta}), \tag{8.40}$$

respectively. In this way, the neural network itself will be able to perturb the baseline in Figure 8.8 so as to achieve almost any airfoil shape with the conditions in Section 8.3.2.

5. Apply lower and upper bounds

The camber distribution is constrained to lie in the interval $y_c \in [\inf(y_c), \sup(y_c)]$. Similarly, the thickness distribution must lie in $y_t \in [\inf(y_t), \sup(y_t)]$. To deal with such constraints we bound the network outputs in the form

$$y_c(x;\underline{\beta}) = \begin{cases} \inf(y_c), & y_c(x;\underline{\beta}) < \inf(y_c). \\ y_c(x;\underline{\beta}), & \inf(y_c) \leq y_c(x;\underline{\beta}) \leq \sup(y_c). \\ \sup(y_c), & y_c(x;\underline{\beta}) > \sup(y_c). \end{cases} \tag{8.41}$$

$$y_t(x;\underline{\beta}) = \begin{cases} \inf(y_t), & y_t(x;\underline{\beta}) < \inf(y_t). \\ y_t(x;\underline{\beta}), & \inf(y_t) \leq y_t(x;\underline{\beta}) \leq \sup(y_t). \\ \sup(y_t), & y_t(x;\underline{\beta}) > \sup(y_t). \end{cases} \tag{8.42}$$

Here we set the lower and upper bounds of the camber and thickness distributions to $\inf(y_c) = -0.05$, $\sup(y_c) = 0.05$, $\inf(y_t) = 0$ and $\sup(y_t) = m/2$.

In summary, the extended class of multilayer perceptron constructed in this section spans a family V of parameterized functions $(y_c, y_t)(x;\underline{\beta})$ of dimension $s = 14+2$, where 14 is the number of biases and synaptic weights in the neural network and 2 is the number of independent parameters. The elements of this function space are thus of the form

$$\begin{aligned} (y_c, y_t): \quad \mathbb{R} &\to \mathbb{R}^2 \\ x &\mapsto (y_c, y_t)(x;\underline{\beta}). \end{aligned}$$

Formulation of variational problem

The objective functional, which is to be maximized, is the aerodynamic efficiency, which is defined by the C_l/C_d ratio. Maximization of C_l/C_d is equivalent to minimization of $-C_l/C_d$.

The statement of the optimal airfoil design problem formulated in this section is then to find a function $(y_c, y_t)^*(x; \underline{\beta}^*)$ for which the functional

$$F[(y_c, y_t)(x; \underline{\beta})] = \frac{C_l}{C_d}, \tag{8.43}$$

defined on V, takes on a minimum value.

In order to evaluate this objective functional is necessary to perform a computational fluid dynamic analysis for each given design. This will provide us with both the lift and drag coefficients. The fluid here is assumed to be Eulerian, and the numerical method used is the finite element method. In order to integrate the Euler equations we use the computational fluid dynamics solver PUMI [79]. Here the computational cost and the quality of the solution are very much dependent on the quality of the mesh used. The software tool used for mesh generation is GiD [25].

Solution of reduced function optimization problem

The quasi-Newton was tested but did not always converge. This may have been due to some noise, which creates many local minima that could 'trap' the quasi-Newton method.

The evolutionary algorithm is applied for training the neural network. Table 8.3.3 provides a summary of the operators and parameters chosen for the evolutionary algorithm. Note that the total number of evaluations here is the population size multiplied by the number of free parameters.

Training Operators	
Fitness assignment method	Linear ranking
Selection method	Stochastic universal sampling
Recombination method	Intermediate
Mutation method	Normal
Training Parameters	
Population size	100
Number of free parameters	16
Selective pressure	1.5
Recombination size	0.25
Mutation rate	0.1
Mutation range	0.1

Table 8.2: Training operators and parameters for the minimum Eulerian drag problem.

These training operators and parameters are chosen due to their statistically reliable performance [81]. The number of free parameters determines the dimensions of the search space and is determined by the multilayer perceptron network architecture. More neurons in the multilayer perceptron will have more parameters, which will be able to represent a greater design complexity. However, as previously discussed, this comes at the cost of computational time. The choice of the population size is a similar compromise. Each individual describes one complete shape and must be evaluated using the fluid solver. But if there are not enough individuals in a population, the algorithm may not explore enough of the solution space to find the global optimum.

It is possible to use already well performing aerofoils to limit the search space, which would reduce the computational time significantly. However, this tends to result in very small modifications to known standard designs. By searching a large function space there is more scope to

discover completely novel designs which would otherwise not be considered. Here the individuals in the population are initialized with random values chosen from a normal distribution with mean 0 and standard deviation 1.

The evaluation history by the evolutionary algorithm is depicted in Figure 8.9. From this picture we can see that the aerodynamic efficiency increases from 4.5 to 4.9, but full convergence is not still achieved, since the evaluation value does not reach a stationary state. This is due to the small number of training generations applied, which is not set to a bigger value due to time constraints. Indeed, training of this neural network lasted one week. Therefore, some form of parallel computing is needed.

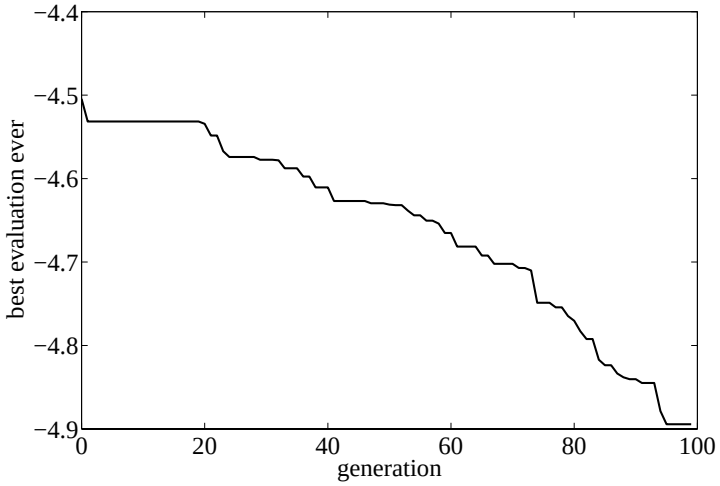


Figure 8.9: Evaluation history in the optimal airfoil design problem.

Table 8.3 shows the training results for this problem. Here M the number of objective function evaluations, CPU the CPU time for a laptop AMD 3000, $\|\underline{\beta}^*\|$ the final parameters norm, $f^*(\underline{\beta}^*)$ the final value for the objective function, p^* the optimal position of maximum thickness, α^* the optimal angle of attack, C_l the corresponding lift coefficient, C_d the corresponding drag coefficient, $\rho(f)$ the mean evaluation of the final population and $\sigma(f)$ the standard deviation of the final population.

The optimal camber and thickness functions represented by the neural network are plotted in Figures 8.10 and 8.11, respectively.

The upper and lower coordinates of this airfoil are depicted in Figure 8.12. This is the shape of a supercritical airfoil.

The pressure coefficient field distribution on the fluid domain is plotted in Figure 8.13.

8.3.4 Conclusions

A class of neural network extended with independent parameters, boundary conditions and lower and upper bounds is able to represent a very complete family of airfoils. This neural network has been used to find an optimal design and its associated angle of attack with given thickness and under transonic flight conditions.

M	=	160000
CPU	=	615276
$\|\underline{\beta}^*\|$	=	1.04762
$f(\underline{\beta}^*)$	=	-4.89434
p^*	=	0.337642
α^*	=	3.890
$C_L(f(\underline{\beta}^*))$	=	0.7613790
$C_D(f(\underline{\beta}^*))$	=	0.1556520
$\rho(f)$	=	-4.79198
$\sigma(f)$	=	0.0366308

Table 8.3: Training results for the optimal airfoil design problem.

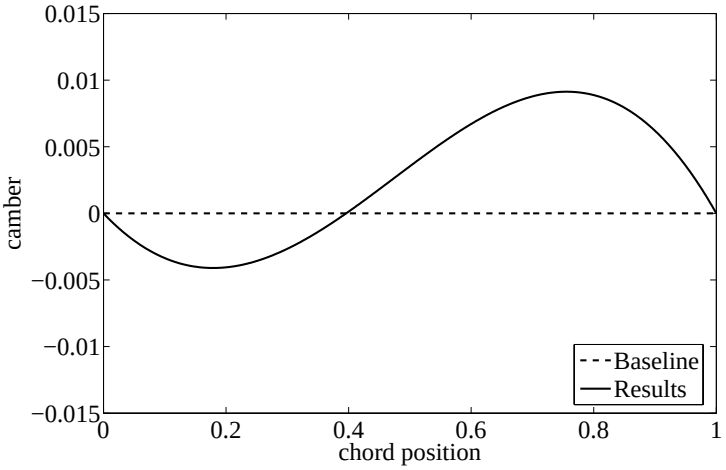


Figure 8.10: Results for the camber in the optimal airfoil design problem.

Future work relies on distributing the computation over a cluster, in order to decrease computational time and to achieve better training results.

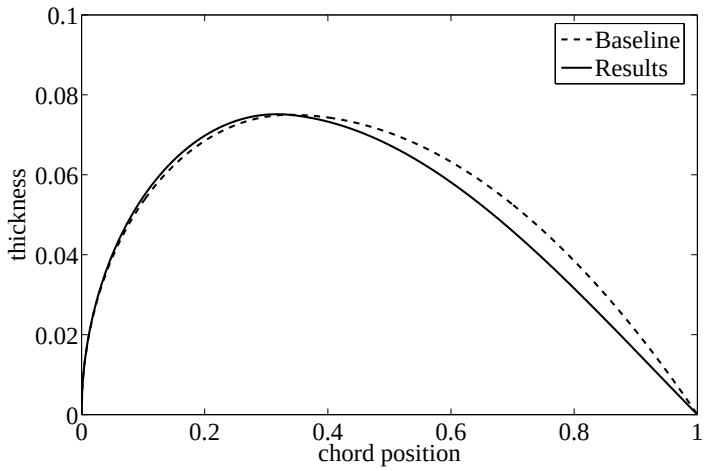


Figure 8.11: Results for the thickness in the optimal airfoil design problem.

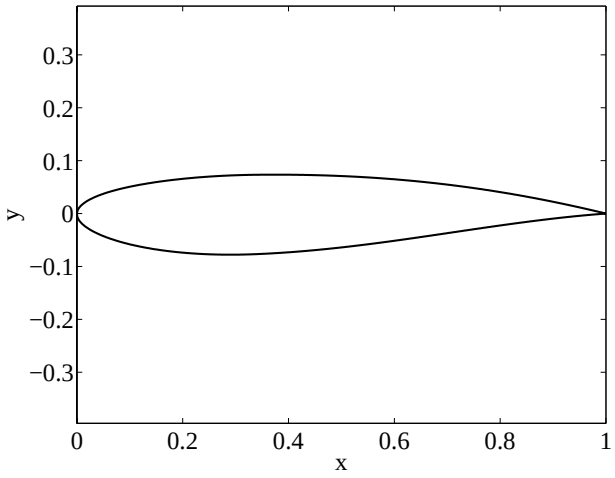


Figure 8.12: Results for the upper and lower surface coordinates in the optimal airfoil design problem.

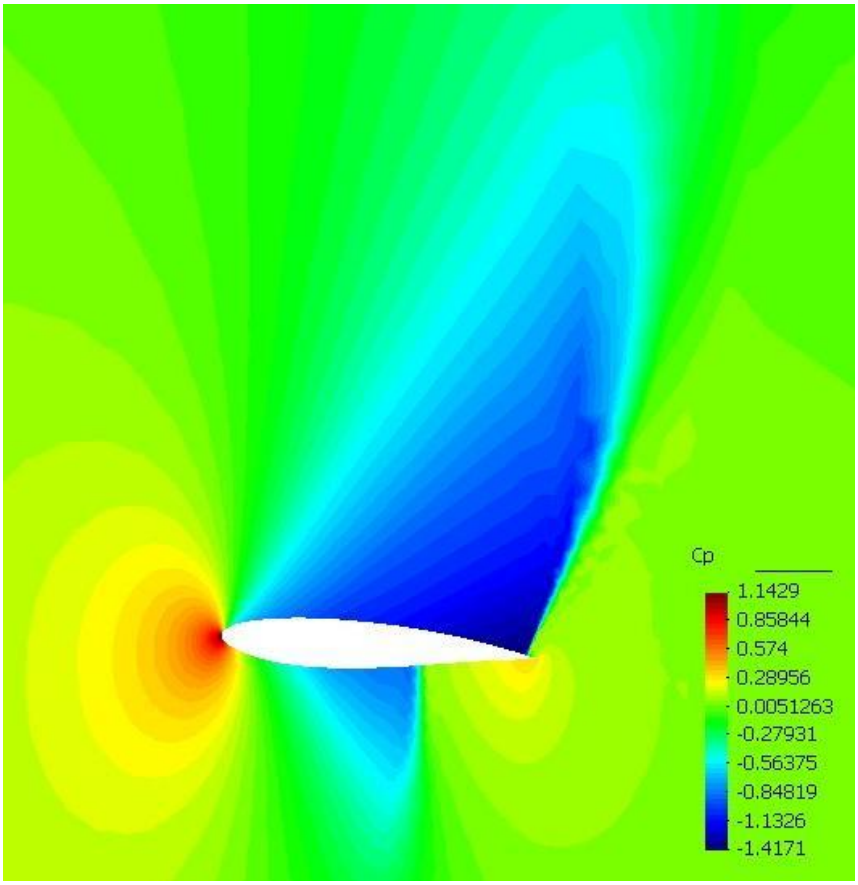


Figure 8.13: Pressure coefficient distribution for the optimal airfoil design.

Chapter 9

Conclusions and future work

The learning problem in the multilayer perceptron has been stated from the point of view of functional analysis and variational calculus. This provides a direct method for the solution of variational problems. The solving approach here consists in three steps: The first step is to choose a suitable parameterized function space in which the solution to the problem is to be approximated. The elements of this family of functions are those spanned by a multilayer perceptron. In the second step the variational problem is formulated by selecting an appropriate objective functional, defined on the function space chosen before. The third step is to solve the reduced function optimization problem. This is performed with a training algorithm capable of finding an optimal set of parameters.

In some cases a standard multilayer perceptron will define a correct representation for the solution. In some other occasions some extensions to this class of neural network shall be required. Here a class of multilayer perceptron extended with independent parameters, boundary conditions and lower and upper bounds has been developed. This extended class of multilayer perceptron might be able to span a more suited function space for some variational problems, so as to deal with a bigger amount of applications.

A variational formulation for the multilayer perceptron is general enough to include traditional learning tasks such as function regression, pattern recognition or time series prediction. In this work these problems have been explained from a variational point of view, and two function regression applications in the naval and the aeronautical industries are solved. The former deals with modeling the residuary resistance of sailing yachts from hull geometry coefficients and Froude number data. In the later the self-noise generated by an airfoil is modeled from a data set of different airfoil geometry coefficients and flight conditions.

The application of this numerical method for problems not involving an input-target data set has been validated through several classical problems in the calculus of variations. In particular, two unconstrained applications, the geodesic and the brachistochrone problems, and two constrained ones, the catenary and the isoperimetric problems, have been solved. The results provided by the neural network have been compared against the analytical ones, demonstrating very good agreement between the approximate and the exact values. In these cases, integration of functions is needed so as to evaluate the objective functional. This is performed with the Runge-Kutta-Fehlberg method.

The optimal control problem has also been formally stated as a learning task for the multilayer perceptron. The car problem, which has analytical solution, has been solved by means of an extended class of multilayer perceptron for validation purposes. Two practical applications in the chemical and the aeronautical industries have also been approached with this numerical method.

In the first one, the feed rate history for a fed batch fermenter providing maximum yield has been found. In the second one, the elevator deflection angle history for an aircraft which gives a safe and comfortable landing has been determined. All these optimal control problems need the integration of a system of ordinary differential equations to evaluate the objective functional, which is also done with the Runge-Kutta-Fehlberg method.

Similarly, the inverse problem has been formulated within a variational theory for the multilayer perceptron. An input estimation and a property estimation problems artificially generated have been approached. The numerical solutions have been compared to the exact ones, with encouraging results. Evaluation of the objective functional for these two problems demands the integration of a partial differential equation. Here the finite element method has been applied for that purpose. On the other hand, an engineering application in the metallurgical industry has also been solved. This problem deals with finding the effective activation energy and the dissolution rate of hardening precipitates for aluminium alloys in thermal treatments. The objective functional for this problem is simply evaluated by solving a set of algebraic equations.

The optimal shape design problem has also been stated as a learning task for the multilayer perceptron. A validation example with analytical solution has again been provided, and the results here show a good suitability of this numerical method for this class of problems. In this case, the integration of a function with the Runge-Kutta-Fehlberg method must be performed to evaluate the objective functional. An engineering application in the aeronautical industry has also been included. Here an airfoil with given thickness has been designed to provide maximum aerodynamic efficiency for transonic flight conditions. The objective functional is evaluated here by integrating a partial differential equation. This is done with the finite element method.

Finally, a software model for the multilayer perceptron is next constructed following a top-down development. The whole process has been carried out in the Unified Modeling Language (UML) and the final implementation has been written in the C++ Programming Language. The result is a comprehensive C++ class library for the solution of function regression, pattern recognition, time series prediction, optimal control, inverse and optimal shape design problems, called Flood and released under an open source license.

In summary, all classes of problems considered in this PhD Thesis have been stated in the same manner, which demonstrates the simplicity of problem formulation. Moreover, every application has been solved up to a great degree of accuracy, showing the good approximation properties of this numerical method. Finally, most of the problems have been successfully approached with first-order training algorithms, which proves the very satisfactory convergence properties of the presented technique.

Ongoing work focuses on the solution of more complex variational problems in engineering by means of a variational formulation for an extended class of multilayer perceptron. Some examples could include many input and output variables, independent parameters, boundary conditions, lower and upper bounds, constraints, partial differential equations and convergence to local optima.

As many of these applications are very expensive in terms of computational demand, it is planned to distribute the different objective functional evaluations over a cluster. This kind of parallelization will drastically reduce the calculation time. Even more, it will make it possible to approach some problems which would be untractable otherwise.

Appendix A

The software model of Flood

Neural networks have been described as a practical tool for the solution of variational problems. This includes data modeling, optimal control, inverse analysis and optimal shape design.

In this Appendix we model a software implementation for the multilayer perceptron. The whole process is carried out in the Unified Modeling Language (UML), which provides a formal framework for the modeling of software systems. The final implementation is to be written in the C++ Programming Language.

A small part of this work has been published in [67].

A.1 The Unified Modeling Language (UML)

The Unified Modeling Language (UML) is a general purpose visual modeling language that is used to specify, visualize, construct, and document the artifacts of a software system [88].

UML class diagrams are the mainstay of object-oriented analysis and design. They show the classes of the system, their interrelationships and the attributes and operations of the classes.

In order to construct a model for the multilayer perceptron, we follow a top-down development. This approach to the problem begins at the highest conceptual level and works down to the details. In this way, to create and evolve a conceptual class diagram for the multilayer perceptron, we iteratively model:

1. Classes.
2. Associations.
3. Derived classes.
4. Attributes and operations.

A.2 Classes

In colloquial terms a concept is an idea or a thing. In object-oriented modeling concepts are represented by means of classes [94]. Therefore, a prime task is to identify the main concepts (or classes) of the problem domain. In UML class diagrams, classes are depicted as boxes [88].

Through all this work, we have seen that neural networks are characterized by a neuron model, a network architecture, an objective functional and a training algorithm. The characterization in classes of these four concepts for the multilayer perceptron is as follows:

Neuron model The class which represents the concept of perceptron neuron model is called `Perceptron`.

Network architecture The class representing the concept of network architecture in the multilayer perceptron is called `MultilayerPerceptron`.

Objective functional The class which represents the concept of objective functional in a multilayer perceptron is called `ObjectiveFunctional`.

Training algorithm The class representing the concept of training algorithm in a multilayer perceptron is called `TrainingAlgorithm`.

Figure A.1 depicts a starting UML class diagram for the conceptual model of the multilayer perceptron.

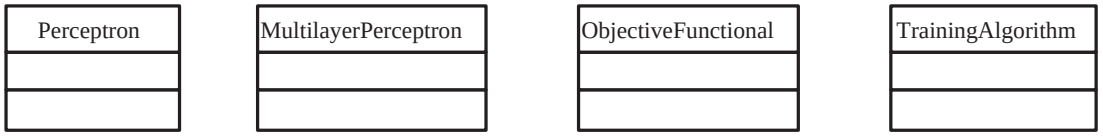


Figure A.1: A conceptual diagram for the multilayer perceptron.

A.3 Associations

Once identified the main concepts in the model it is necessary to aggregate the associations among them. An association is a relationship between two concepts which points some significative or interesting information [94]. In UML class diagrams, an association is shown as a line connecting two classes. It is also possible to assign a label to an association. The label is typically one or two words describing the association [88].

The appropriate associations in the system are next identified to be included to the UML class diagram of the system:

Neuron model - Multilayer perceptron A multilayer perceptron *is built by* perceptrons.

Network architecture - Objective functional A multilayer perceptron *has assigned* an objective functional.

Objective functional - Training algorithm An objective functional *is improved by* a training algorithm.

Figure A.2 shows the above UML class diagram with these associations aggregated.

A.4 Derived classes

In object-oriented programming, some classes are designed only as a parent from which sub-classes may be derived, but which is not itself suitable for instantiation. This is said to be an *abstract class*, as opposed to a *concrete class*, which is suitable to be instantiated. The derived class contains all

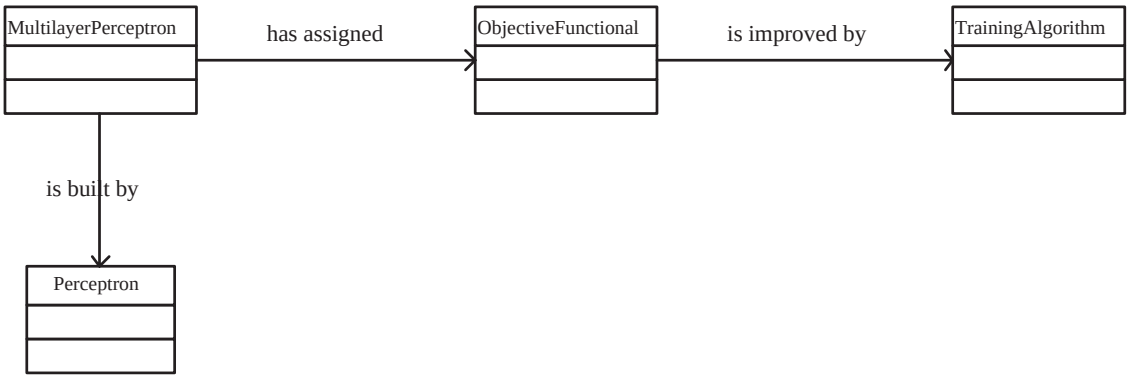


Figure A.2: Aggregation of associations to the conceptual diagram.

the features of the base class, but may have new features added or redefine existing features [94]. Associations between a base class an a derived class are of the kind *is a* [88].

The next task is then to establish which classes are abstract and to derive the necessary concrete classes to be added to the system. Let us then examine the classes we have so far:

Neuron model The class `Perceptron` is abstract, because it does not represent any concrete neuron model, since it must be assigned a specific activation function.

On the other hand, a multilayer perceptron with a sigmoid hidden layer and a linear output layer meets the universal approximation theorem. Therefore, concrete classes for the sigmoid perceptron and the linear perceptron must be derived. They are called `SigmoidPerceptron` and `LinearPerceptron`, respectively.

Network architecture The class `MultilayerPerceptron` is a concrete class and is itself suitable for instantiation.

Objective functional The class `ObjectiveFunctional` is abstract, because it does not represent a concrete objective functional for the multilayer perceptron.

For function regression problems, the sum of squares error can approximate the conditional distribution of the output variables, conditioned on the input variables. For pattern recognition problems, the sum of squares error can also approximate the posterior probabilities of class membership, again conditioned on the input variables. Therefore we derive a class for the sum of squares error, which is called `SumOfSquaresError`. There might arise some occasions when we want to regularize the neural network's output. For that reason we add the class `SumOfSquaresErrorWithFreeParameterDecay` to the model. Both the sum of squares error and the sum of squares error with free parameter decay objective functionals are measured on a training data set, so we add to the model a class which represents that concept. This is called `TrainingDataSet`, and it is a concrete class.

In order to solve the brachistochrone and the isoperimetric problems the classes `BrachistochroneProblem` and `IsoperimetricProblem` are derived, respectively. To evaluate the performance of the neural network in that two problems we make use of numerical integration. In this way, we add to the model an utility class called `IntegrationOfFunctions`.

It is always possible to derive new objective functionals for the multilayer perceptron at any time and include them in the system.

Training algorithm The class `TrainingAlgorithm` is abstract, because it does not represent a training algorithm for an objective function of a multilayer perceptron.

Several training algoritm classes have been implemented. These include `GradientDescent`, `QuasiNewtonMethod` or `EvolutionaryAlgorithm`, for instance.

Figure A.3 shows the UML class diagram for the multilayer perceptron with some of the derived classes included.

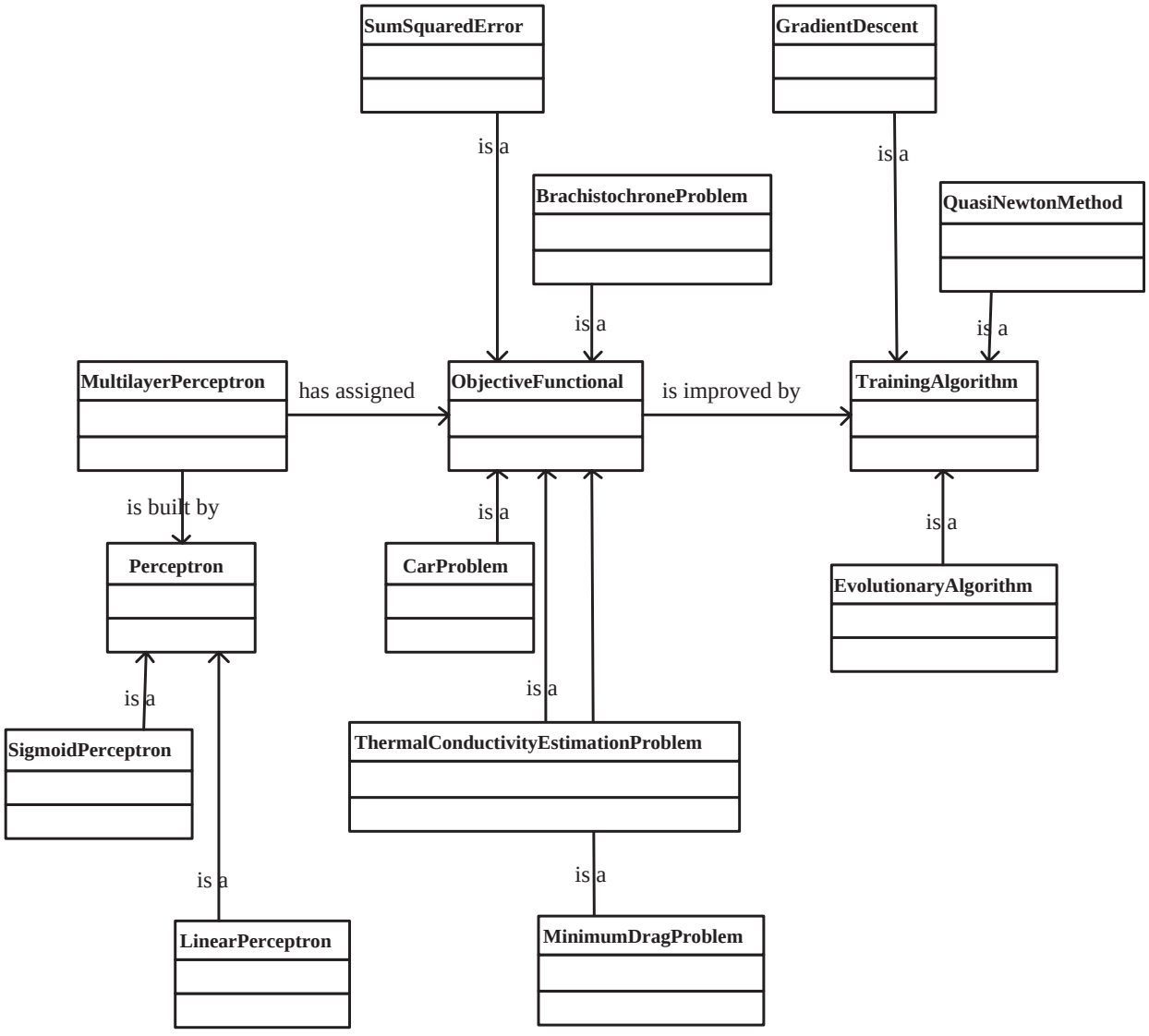


Figure A.3: Aggregation of derived classes to the association diagram.

A.5 Attributes and operations

An attribute is a named value or relationship that exists for all or some instances of a class. An operation is a procedure associated with a class [94]. In UML class diagrams, classes are depicted as boxes with three sections: the top one indicates the name of the class, the one in the middle lists the attributes of the class, and the bottom one lists the operations [88].

A.5.1 Perceptron

A perceptron neuron model has the following attributes:

1. A number of neuron inputs.
2. A set of synaptic weights.
3. A bias.

and performs the following operations:

1. Get the output value for a given set of input signals.
2. Get the output derivative value for a given set of input signals.

Figure A.4 shows the class `Perceptron` together with its derived classes, `SigmoidPerceptron` and `LinearPerceptron`. It includes the attributes and the operations of the base class and the derived classes.

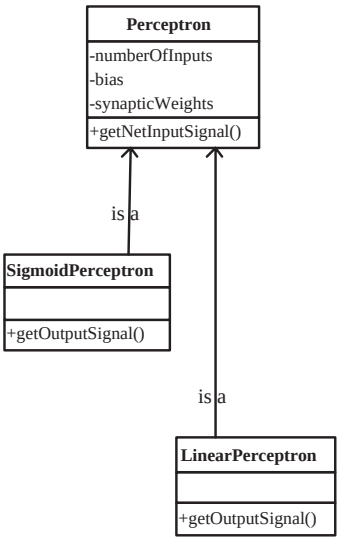


Figure A.4: Attributes and operations of the Perceptron classes.

A.5.2 Multilayer perceptron

A multilayer perceptron has the following attributes:

1. A number of network inputs.
2. A hidden layer of a given number of sigmoid neurons.
3. An output layer of a given number of linear neurons.

and performs the following operations:

1. Get the set of output signals for a given set of input signals.
1. Get the Jacobian matrix for a given set of input signals.

Figure A.5 shows the attributes and operations of the class MultilayerPerceptron.

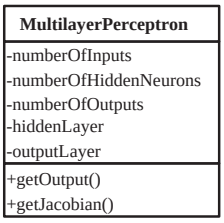


Figure A.5: Attributes and operations of the class MultilayerPerceptron.

A.5.3 Objective functional

An objective functional for a multilayer perceptron has the following attributes:

1. A relationship to a multilayer perceptron. In C++ this is implemented as a pointer to a multilayer perceptron object.

and performs the following operations:

1. Get the evaluation of a multilayer perceptron.
2. Get the objective function gradient vector of a multilayer perceptron.

On the other hand, a training data set has the following attributes:

1. A number of samples.
2. A number of input variables.
3. A number of target variables.
4. A set of input data.
5. A set of target data.

and performs the following operations:

1. Load the training data set from a data file.

Last, the utility class for numerical integration has no attributes, and performs a unique operation:

1. Get the definite integral of a function.

Figure A.6 shows the class `ObjectiveFunctional` together with some of its derived classes. It includes the attributes and the operations of the base class and the derived classes.

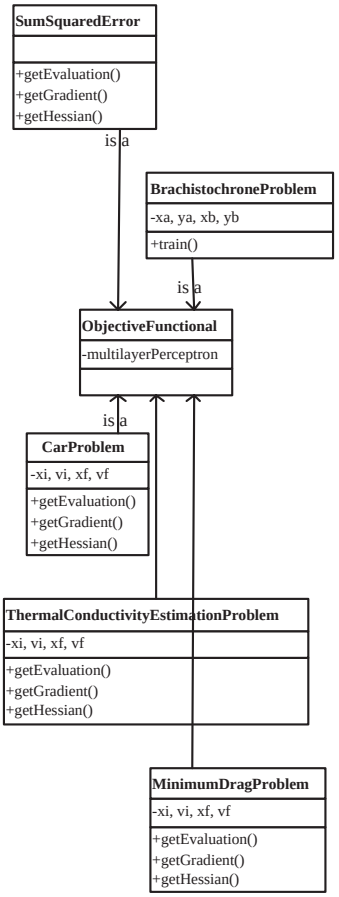


Figure A.6: Attributes and operations of some ObjectiveFunctional classes.

A.5.4 Training algorithm

A training algorithm for a multilayer perceptron has the following attributes:

1. A relationship to an objective functional for a multilayer perceptron. In C++ this is implemented as a pointer to an objective functional object.
2. A set of training parameters.

and performs the following operations:

1. Train a multilayer perceptron.

Figure A.7 shows the class `TrainingAlgorithm` together with its derived classes. It includes the attributes and the operations of the base class and the derived classes.

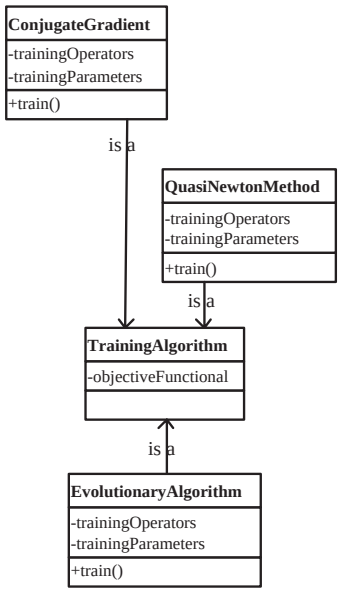


Figure A.7: Attributes and operations of the TrainingAlgorithm classes.

Appendix B

Numerical integration

Evaluation of the objective functional of a neural network often requires to integrate functions, ordinary differential equations and partial differential equations. In general, approximate approaches need to be applied here.

Some standard methods for numerical integration are described in this Appendix. They are meant to be utilities which can be embedded into a specific problem when necessary.

B.1 Integration of functions

B.1.1 Introduction

Numerical integration is the approximate computation of an integral using numerical techniques. More specifically, the problem is to compute the definite integral of a given real function $f(x)$ over a closed interval $[a, b]$,

$$I[y(x)] = \int_a^b f(x)dx. \tag{B.1}$$

There are a wide range of methods available for numerical integration [83]. The numerical computation of an integral is sometimes called quadrature.

B.1.2 Closed Newton-Cotes formulas

The Newton-Cotes formulas are an extremely useful and straightforward family of numerical integration techniques. The integration formulas of Newton and Cotes are obtained dividing the interval $[a, b]$ into n equal parts such that $f_n = f(x_n)$ and $h = (b - a)/n$. Then the integrand $f(x)$ is replaced by a suitable interpolating polynomial $P(x)$, so that

$$\int_a^b f(x) \sim \int_a^b P(x). \tag{B.2}$$

To find the fitting polynomials, the Newton-Cotes formulas use Lagrange interpolating polynomials [93]. Next we examine some rules of this kind.

The trapezoidal rule

The 2-point closed Newton-Cotes formula is called the trapezoidal rule, because it approximates the area under a curve by a trapezoid with horizontal base and sloped top (connecting the endpoints a and b). In particular, let call the lower and upper integration limits x_0 and x_1 respectively, the integration interval h, and denote $f_n = f(x_n)$. Then the trapezoidal rule states that [99]

$$\begin{aligned} \int_{x_1}^{x_2} f(x)dx &= h\left[\frac{1}{2}f_1 + \frac{1}{2}f_2\right] \\ &+ \mathcal{O}(h^3 f''), \end{aligned} \tag{B.3}$$

were the error term $O(\cdot)$ means that the true answer differs from the estimate by an amount that is the product of some numerical coefficient times h^3 times the value of the functions second derivative somewhere in the interval of integration.

Simpson's rule

The 3-point closed Newton-Cotes formula is called Simpson's rule. It approximates the integral of a function using a quadratic polynomial. In particular, let the function f be tabulated at points x_0, x_1, and x_2, equally spaced by distance h, and denote $f_n = f(x_n)$. Then Simpson's rule states that [99]

$$\begin{aligned} \int_{x_1}^{x_3} f(x)dx &= h\left[\frac{1}{3}f_1 + \frac{4}{3}f_2 + \frac{1}{3}f_3\right] \\ &+ \mathcal{O}(h^5 f^{(4)}). \end{aligned} \tag{B.4}$$

Here $f^{(4)}$ means the fourth derivative of the function f evaluated at an unknown place in the interval. Note also that the formula gives the integral over an interval of size $2h$, so the coefficients add up to 2.

B.1.3 Extended Newton-Cotes formulas

The Newton-Cotes formulas are usually not applied to the entire interval of integration $[a, b]$, but are instead used in each one of a collection of subintervals into which the interval $[a, b]$ has been divided. The full integral is the approximated by the sum of the approximations to the subintegrals. The locally used integration rule is said to have been extended, giving rise to a composite rule [93]. We proceed to examine some composite rules of this kind.

Extended trapezoidal rule

For n tabulated points, using the trapezoidal rule $n - 1$ times and adding the results gives [99]

$$\begin{aligned} \int_{x_1}^{x_n} f(x)dx &= h\left[\frac{1}{2}f_1 + f_2 + f_{n-1} + \frac{1}{2}f_n\right] \\ &+ \mathcal{O}\left(\frac{(b-a)^3 f''}{N^2}\right). \end{aligned} \tag{B.5}$$

Note that the error estimate is here written in terms of the interval $b - a$ and the number of points N instead of in terms of h.

Extended Simpson's rule

For and odd number n of tabulated points, the extended Simpson's rule is [99]

$$\begin{aligned}\int_{x_1}^{x_n} f(x)dx &= h\left[\frac{1}{3}f_1 + \frac{4}{3}f_2 + \frac{2}{3}f_3 + \frac{4}{3}f_4 + \ldots + \frac{2}{3}f_{n-2} + \frac{4}{3}f_{n-1} + \frac{1}{3}f_n\right] \\ &+ \mathcal{O}\left(\frac{1}{N^4}\right). \end{aligned} \tag{B.6}$$

B.1.4 Ordinary differential equation approach

The evaluation of the integral (B.1) is equivalent to solving for the value $I \equiv y(b)$ the ordinary differential equation

$$\frac{dy}{dx} = f(x), \tag{B.7}$$

$$y(a) = 0. \tag{B.8}$$

Section B.2 of this PhD Thesis deals with the numerical integration of differential equations. In that section, much emphasis is given to the concept of 'variable' or 'adaptive' choices of stepsize.

B.2 Ordinary differential equations

B.2.1 Introduction

An ordinary differential equation (ODE) is an equality involving a function and its derivatives. An ODE of order n is an equation of the form

$$F(x, y(x), y'(x), \ldots, y^{(n)}(x)) = 0, \tag{B.9}$$

where y is a function of x, $y' = dy/dx$ is the first derivative of y with respect to x, and $y^{(n)} = dy^n/dx^n$ is the n-th derivative of y with respect to x.

The generic problem of a n-th order ODE can be reduced to the study of a set of n coupled first-order differential equations for the functions y_i, $i = 1, \ldots, n$, having the general form

$$\frac{dy_i(x)}{dx} = f_i(x, y_1, \ldots, y_N), \tag{B.10}$$

for $i = 1, \ldots, n$ and where the functions f_i on the right-hand side are known.

While there are many general techniques for analytically solving different classes of ODEs, the only practical solution technique for complicated equations is to use numerical methods. The most popular of these are the Runge-Kutta and the Runge-Kutta-Fehlberg methods.

A problem involving ODEs is not completely specified by its equations. In initial value problems all the y_i are given at some starting value x's, and it is desired to find the y_i's at some final point x_f, or at some discrete list of points (for example, at tabulated intervals).

B.2.2 The Euler method

The formula for the Euler method is

$$y_{n+1} = y_n + hf(x_n, y_n) + \mathcal{O}(h^2), \tag{B.11}$$

which advances a solution from x_n to $x_{n+1} = x_n + h$. The formula is unsymmetrical: It advances the solution through an interval h, but uses derivative information only at the beginning of that interval.

There are several reasons that Eulers method is not recommended for practical use, among them, (i) the method is not very accurate when compared to other methods run at the equivalent stepsize, and (ii) neither is it very stable [83].

B.2.3 The Runge-Kutta method

Consider the use of a step like (B.11) to take a 'trial' step to the midpoint of the interval. The use the value of both x and y at that midpoint to compute the 'real' step across the whole interval. This can be written

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + \frac{1}{2}h, y_n + \frac{1}{2}k_1) \\ y_{n+1} &= y_n + k_2 + \mathcal{O}(h^3) \end{aligned} \tag{B.12}$$

As indicated in the error term, this symmetrization cancels out the first-order error term, making the method second order. A method is conventionally called nth order if its error term is $\mathcal{O}(h^{n+1})$- In fact Equation (B.12) is called the second-order Runge-Kutta or midpoint method.

There are various specific formulas that derive from this basic idea. The most often used is the classical fourth-order Runge-Kutta formula,

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + \frac{h}{2}, y_n + \frac{k_1}{2}) \\ k_3 &= hf(x_n + \frac{h}{2}, y_n + \frac{k_2}{2}) \\ k_4 &= hf(x_n + h, y_n + k_3) \\ y_{n+1} &= y_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6} + \mathcal{O}(h^5) \end{aligned} \tag{B.13}$$

The fourth order Runge-Kutta method requires four evaluations of the right-hand side per step h.

B.2.4 The Runge-Kutta-Fehlberg method

Implementation of adaptive stepsize control requires that the stepping algorithm signal information about its performance, most important, an estimate of its truncation error.

It is this difference that we shall endeavor to keep to a desired degree of accuracy, neither too large nor too small. We do this by adjusting h.

An step size adjustment algorithm is based on the embedded Runge-Kuta formulas, originally invented by Fehlberg. The general form of a fifth-order Runge-Kutta formula is

$$\begin{aligned} k_1 &= hf(x_n, y_n) \\ k_2 &= hf(x_n + a_2 h, y_n + b_{21} k_1) \\ &\dots \\ k_6 &= hf(x_n + a_6 h, y_n + b_{61} k_1 + \ldots + b_{65} k_5) \\ y_{n+1} &= y_n + c_1 k_1 + c_2 k_2 + c_3 k_3 + c_4 k_4 + c_5 k_5 + c_6 k_6 + \mathcal{O}(h^6) \end{aligned} \tag{B.14}$$

The embedded fourth order formula is

$$y^*_{n+1} = y_n + c^*_1 k_1 + c^*_2 k_2 + c^*_3 k_3 + c^*_4 k_4 + c^*_5 k_5 + c^*_6 k_6 + \mathcal{O}(h^5) \tag{B.15}$$

And so the error estimate is

$$\begin{aligned} \Delta &\equiv y_{n+1} - y^*_{n+1} \\ &= \sum_{i=1}^{6} (c_i - c^*_i) k_i. \end{aligned} \tag{B.16}$$

The particular values of the various constants that we favor are those found by Cash and Karp [21], and given in Table B.1. These give a more efficient method than Fehlberg's original values [83].

i	a_i	b_{i1}	b_{i2}	b_{i3}	b_{i4}	b_{i5}	c_i	c^*_i
1							$\frac{37}{378}$	$\frac{2825}{27648}$
2	$\frac{1}{5}$	$\frac{1}{5}$					0	0
3	$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$				$\frac{250}{621}$	$\frac{18575}{48384}$
4	$\frac{3}{5}$	$\frac{3}{10}$	$-\frac{9}{10}$	$\frac{6}{5}$			$\frac{125}{594}$	$\frac{13525}{55296}$
5	1	$-\frac{11}{54}$	$\frac{5}{2}$	$-\frac{70}{27}$	$\frac{35}{27}$		0	$\frac{277}{14336}$
6	$\frac{7}{8}$	$\frac{1631}{55296}$	$\frac{175}{512}$	$\frac{575}{13824}$	$\frac{44275}{110592}$	$\frac{253}{4096}$	$\frac{512}{1771}$	$\frac{1}{4}$

Table B.1: Cash and Karp parameters for the Runge-Kutta-Fehlberg method.

Now that we know, approximately, what the error is, we need to consider how to keep it within desired bounds. What is the relation between Δ and h? According to Equations (B.14) and (B.15), Δ scales as h^5. If we take a step h_1 and produce an error Δ_1, therefore, the step h_0 that would have given some other value Δ_0 is readily estimated as

$$h_0 = h_1 \left| \frac{\Delta_0}{\Delta_1} \right|^{0.2} \tag{B.17}$$

Henceforth we will let Δ_0 denote the desired accuracy. Then, Equation (B.17) is used in two ways: If Δ_1 is larger than Δ_0 in magnitude, the equation tells how much to decrease the step when

we retry the present (failed) step. If Δ_1 is smaller than Δ_0, on the other hand, then the equation tells how much we can safely increase the stepsize for the next step.

This notation hides the fact that Δ_0 is actually a vector of desired accuracies, one for each equation in the set of ODEs. In general, the accuracy requirement will be that all equations are within their respective allowed errors.

$$h_0 = \begin{cases} Sh_1 \left|\frac{\Delta_0}{\Delta_1}\right|^{0.20} & \Delta_0 \geq \Delta_1 \\ Sh_1 \left|\frac{\Delta_0}{\Delta_1}\right|^{0.25} & \Delta_0 < \Delta_1 \end{cases} \tag{B.18}$$

B.3 Partial differential equations

B.3.1 Introduction

Partial differential equations arise in all fields of science and engineering, since most real physical processes are governed by them. A partial differential equation is an equation stating a relationship between a function of two or more independent variables and the partial derivatives of this function with respect to the independent variables. In most problems, the independent variables are either space (x, y, z) or space and time (x, y, z, t). The dependent variable depends on the physical problem being modeled.

Partial differential equations are usually classified into the three categories, hyperbolic, parabolic, and elliptic, on the basis of their characteristics [83].

The prototypical example of a hyperbolic equation is the wave equation

$$\frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = v^2 \nabla^2 u(\mathbf{x}, t), \tag{B.19}$$

where v is the velocity of wave propagation.

The prototypical parabolic equation is the diffusion equation

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = \nabla(D \nabla u(\mathbf{x}, t)), \tag{B.20}$$

where D is the diffusion coefficient.

The prototypical elliptic equation is the Poisson equation

$$\nabla^2 u(\mathbf{x}) = \rho(\mathbf{x}), \tag{B.21}$$

where the source term ρ is given. If the source term is equal to zero, the equation is Laplace's equation.

From a computational point of view, the classification into these three canonical types is not as important as some other essential distinctions [83]. Equations (B.19) and (B.20) both define initial value (or Cauchy) problems. By contrast, Equation (B.21) defines a boundary value problem.

In an initial value problem information of u is given at some initial time t_0 for all $\mathbf{x}$. The PDE then describes how $u(\mathbf{x}, t)$ propagates itself forward in time [83].

On the other hand, boundary value problems direct to find a single static function u which satisfies the equation within some region of interest $\mathbf{x}$, and which has some desired behavior on the boundary of that region [83].

In a very few special cases, the solution of a PDE can be expressed in closed form. In the majority of problems in engineering and science, the solution must be obtained by numerical methods.

B.3.2 The finite differences method

In the finite differences approach, all the derivatives in a differential equation are replaced by algebraic finite difference approximations, which changes the differential equation into an algebraic equation that can be solved by simple arithmetic [45].

The error between the approximate solution and the true solution here is determined by the error that is made by going from a differential operator to a difference operator. This error is called the discretization error or truncation error [76]. The term truncation error reflects the fact that a difference operator can be viewed as a finite part of the infinite Taylor series of the differential operator.

B.3.3 The finite element method

Another approach for solving differential equations is based on approximating the exact solution by an approximate solution, which is a linear combination of specific trial functions, which are typically polynomials. These trial functions are linearly independent functions that satisfy the boundary conditions. The unknown coefficients in the trial functions are then determined by solving a system of linear algebraic equations [45].

Because finite element methods can be adapted to problems of great complexity and unusual geometries, they are an extremely powerful tool in the solution of important problems in science and engineering. It is out of the scope of this work to provide a detailed explanation of the finite element method. The interested reader is referred to [102].

This page is intentionally left blank.

Appendix C

Related publications

Here a list of publications resulted from this PhD Thesis is written.

Publications in international journals

R. Lopez, E. Balsa-Canto and E. Oñate. *Neural networks for variational problems in engineering*, International Journal for Numerical Methods in Engineering, Volume 75 Issue 11, Pages 1341-1360, 2008.

R. Lopez and E. Oñate. *An extended class of multilayer perceptron*, Neurocomputing, Volume 71, Issues 13-15, Pages 2538-2543, 2008.

Publications in international conference proceedings

R. Lopez, R. Flores, E. Escolano and E. Oñate. *Neural networks for optimal airfoil design*, in Proceedings of the Third International Conference on Advances and Applications of GiD GiD2008, 2008.

E. Oñate, J. Piazzese, R. Lopez, G. Corestein, E. Blade, M. Gomez-Valentin and J. Dolz. *RAMFLOOD decision support system based on artificial neural networks for risk assessment and management of floods*, in Proceedings of the European Conference on Flood Risk Management Research into Practice FLOODrisk 2008, 2008.

R. Lopez, X. Diego, R. Flores, M. Chiumenti, and E. Oñate. *Artificial neural networks for the solution of optimal shape design problems*, in Proceedings the 8th World Congress on Computational Mechanics WCCM8, 2008.

R. Lopez, B. Ducoeur, M. Chiumenti, B. de Meester and C. Agelet de Saracibar. *Modeling precipitation dissolution in hardened aluminium alloys using neural networks*, in Proceedings of the 11th International Conference on Material Forming ESAFORM 2008, 2008.

I. Ortigosa, R. Lopez and J. Garcia. *A neural networks approach for prediction of total resistance's coefficients*, in Proceedings of the Fifth International Congress on Maritime Technological Innovations and Research ICMTIR 2007, 2007.

E. Soudah, J.F. Rodriguez, R. Lopez and E. Oñate. *Neural networks for simulation of computational processes on bioengineering*, in Proceedings of the Third World Congress on Bioengineering WACBE 2007, 2007.

K. Lau, R. Lopez and E. Oñate. *Neural networks for optimal control of aircraft landing systems*, in Proceedings of the World Congress on Engineering WORLDENG-ICAEM 2007, 2007.

I. Ortigosa, R. Lopez and J. Garcia. *A neural networks approach to residuary resistance of sailing yachts prediction*, in Proceedings of the International Conference on Marine Engineering MARINE 2007, 2007.

R. Lopez and E. Oñate. *A software model for the multilayer perceptron*, in Proceedings of the International Conference on Applied Computing IADIS 2007, 2007.

R. Lopez and E. Oñate. *A variational formulation for the multilayer perceptron*, in Proceedings of the 16th International Conference on Artificial Neural Networks ICANN 2006, 2006.

P. Dadvand, R. Lopez and E. Oñate. *Artificial neural networks for the solution of inverse problems*, in Proceedings of the International Conference on Design Optimization Methods and Applications ERCOFTAC 2006, 2006.

R. Lopez, E. Balsa-Canto and E. Oñate. *Artificial neural networks for the solution of optimal control problems*, in Proceedings of the Sixth International Conference on Evolutionary and Deterministic Methods for Design, Optimisation and Control with Applications to Industrial and Societal Problems EUROGEN 2005, 2005.

Appendix D

Related projects

Here we refer some international research projects in which the contents of this PhD Thesis have been applied.

01/12/2007 - 01/03/2010:
Development of an advanced Hybrid Evolutionary Algorithm for the optimisation of Aeronautical Configurations in a Multi-objective space (DRAGON)
AIRBUS Contract.

01/09/2006 - 29/02/2009:
Risk Assessment Model of the Water-Sediment-Soil (RAMWASS)
European Commission, Ref. GOCE-CT-2006-037081.

01/04/2005 - 31/03/2008:
Detailed Multi-Physics Modelling of Friction Stir Welding (DEEPWELD)
European Commission, Ref. AST4-CT-2005-516134.

01/11/2004 - 31/10/2006:
Grid Based Decision Support System for Clinical Diagnosis and Interventions in Cardiovascular Problems (DISHEART)
European Commission, Ref. COOP-CT-2004-513226.

01/10/2002 - 30/09/2003:
Decision Support System for Risk Assessment and Management of Floods (RAMFLOOD)
European Commission, Ref. IST-2001-37581.

This page is intentionally left blank.

Appendix E

Related software

Here a software package which is a result of this PhD Thesis is cited.

Flood: An Open Source Neural Networks C++ Library (www.cimne.com/flood).

This page is intentionally left blank.

References

[1] U.s. centennial of flight commission. www.centennialofflight.gov, 2006.

[2] R.K. Amiet. Effect of the incident surface pressure field on noise due to a turbulent flow past a trailing edge. *Journal of Sound and Vibration*, 57(2):305–306, 1978.

[3] J.A. Anderson and E. Rosenfeld, editors. *Talking Nets: An Oral History of Neural Networks*. Bradford Books, 1998.

[4] R. Aris. *Elementary Chemical Reactor Analysis*. Butterworths, 1989.

[5] H. Ashley. *Engineering Analysis of Flight Vechicles*. Dover Publishing, 1992.

[6] R. Audi, editor. *The Cambridge Dictionary of Philosophy*. Cambridge University Press, 1999.

[7] A.W. Babister. *Aircraft Dynamic Stability and Response*. Pergamon Press, 1980.

[8] T. Bäck and F. Hoffmeister. Extended selection mechanisms in genetic algorithms. In *Proceedings of the Fourth International Conference on Genetic Algorithms, San Mateo, California, USA*, pages 92–99, 1991.

[9] J.E. Baker. Reducing bias and inefficiency in the selection algorithm. In *Proceedings of the Second International Conference on Genetic Algorithms, Hillsdale, New Jersey, USA*, pages 14–21, 1987.

[10] E. Balsa-Canto. *Algoritmos Eficientes para la Optimizacion Dinamica de Procesos Distribuidos*. PhD thesis, Universidad de Vigo, 2001.

[11] R. Battiti. First and second order methods for learning: Between steepest descent and newton's method. *Neural Computation*, 4(2):141–166, 1992.

[12] E.B. Baum and F. Wilczek. What size net gives valid generalization? *Neural Computation*, 1(1):151–160, 1989.

[13] L. Belanche. *Heterogeneous Neural Networks*. PhD thesis, Technical University of Catalonia, 2000.

[14] J.T. Betts. A survey of numerical methods for trajectory optimization. *AIAA Journal of Guidance, Control and Dynamics*, 21(2):193–207, 1998.

[15] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[16] T.F. Brooks, D.S. Pope, and A.M. Marcolini. Airfoil self-noise and prediction. Technical report, NASA RP-1218, July 1989.

[17] D.G. Brown S.L., Hull. Axisymmetric bodies of minimum drag in hypersonic flow. *Journal of Optimization Theory and Applications*, 3(1):52–71, 1969.

[18] D. Bucur and G. Buttazzo. *Variational Methods in Shape Optimization Problems*. Birkhauser, 2005.

[19] Hwang C. and Y.P. Shih. Laguerre series direct method for variational problems. *Journal of Optimization Theory and Applications*, 39(1):143–149, 1983.

[20] G.A. Carpenter and S. Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37(1):54–115, 1987.

[21] J.R. Cash and A.H. Karp. A variable order runge-kutta method for initial value problems with rapidly varying right hand sides. *ACM Transactions on Mathematical Software*, 16(3):201–222, 1990.

[22] R.Y. Chang and M.L. Wang. Shifted legendre direct method for variational problems. *Journal of Optimization Theory and Applications*, 39(2):299–307, 1983.

[23] C.T. Chen and C. Hwang. Optimal control computation for differential-algebraic process systems with general constraints. *Chemical Engineering Communications*, 97:9–26, 1990.

[24] Z. Chen and S. Haykin. On different facets of regularization theory. *Neural Computation*, 14(12):2791–2846, 2002.

[25] CIMNE. Gid: The personal pre and post processor. www.gidhome.com, 2006.

[26] P. Dadvand, R. Lopez, and E. Oñate. Artificial neural networks for the solution of inverse problems. In *Proceedings of the International Conference on Design Optimisation Methods and Applications ERCOFTAC 2006*, 2006.

[27] H. Demuth and M. Beale. *Neural Network Toolbox for Use with MATLAB. User's Gide*. The MathWorks, 2002.

[28] F.J. Ellert and C.W. Merriam. Synthesis of feedback controls using optimization theory - an example. *IEEE Transactions on Automatic Control*, 8(2):89–103, 1963.

[29] L.E. Elsgolc. *Calculus of Variations*. Pergamon Press, 1961.

[30] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Springer, 2000.

[31] S. Eyi, J.O. Hager, and K.D. Lee. Airfoil design optimization using the navier-stokes equations. *Journal of Optimization Theory and Applications*, 83(3):447–461, 1994.

[32] W.E. Failer and S.J. Schrec. Neural networks: Applications and opportunities in aeronautics. *Progress in Aerospace Sciences*, 32:433–456, 1996.

[33] R. Fletcher and C.M. Reeves. Function minimization by conjugate gradients. *Computer Journal*, 7:149–154, 1964.

[34] D.B. Fogel. An introduction to simulated evolutionary optimization. *IEEE Transactions on Neural Networks*, 5(1):3–14, 1994.

[35] E. Franco-Lara and D. Weuster-Botz. Estimation of optimal feeding strategies for fed-bach bioprocesses. *Bioprocess and Biosystems Engineering*, 27(4):255–262, 2005.

[36] I.M. Gelfand and S.V. Fomin. *Calculus of Variations*. Prentice Hall, 1963.

[37] J. Gerritsma, R. Onnink, and A. Versluis. Geometry, resistance and stability of the delft systematic yacht hull series. In *International Shipbuilding Progress*, volume 28, pages 276–297, 1981.

[38] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural network architectures. *Neural Computation*, 7(2):219–269, 1995.

[39] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, 1988.

[40] M.T. Hagan and M. Menhaj. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks*, 5(6):989–993, 1994.

[41] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1994.

[42] D.O. Hebb. *The Organization of Behaviour*. New York: Wiley, 1949.

[43] Charles Hirsch. *Numerical computation of internal and external flows. Vol. 1 - Fundamentals of Numerical Discretization*. John Wiley & Sons, 1990.

[44] Charles Hirsch. *Numerical computation of internal and external flows. Vol. 2 - Computational Methods for Inviscid and Viscous Flows*. John Wiley & Sons, 1990.

[45] J.D. Hoffman. *Numerical Methods for Engineers and Scientists*. CRC Press, second edition, 2001.

[46] J. Hong. Optimal substrate feeding policy for a fed batch fermentation with substrate and product inhibition kinetics. *Biotechnology and Bioengineering*, 28:14211431, 1986.

[47] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Sciences*, volume 79, pages 2554–2558, 1982.

[48] I.R. Horng and J.H. Chou. Shifted chebyshev direct method for solving variational problems. *International Journal of Systems Science*, 16(77):855–861, 1985.

[49] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[50] M.S. Howe. A review of the theory of trailing edge noise. *Journal of Sound and Vibration*, 61:437–465, 1978.

[51] C.H. Hsiao. Solution of variational problems via haar orthonormal wavelet direct method. *International Journal of Computer Mathematics*, 81(7):871–887, 2004.

[52] J. Kennedy and R.C. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1948, 1995.

[53] D.E. Kirk. *Optimal Control Theory. An Introduction*. Prentice Hall, 1970.

[54] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, 1996.

[55] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[56] A.N. Kolmogorov and Fomin S.V. *Elements of the Theory of Functions and Functional Analysis*. Graylock Press, 1961.

[57] Kevin Lau. A neural networks approach for aerofoil noise prediction. Master's thesis, Department of Aeronautics. Imperial College of Science, Technology and Medicine (London, United Kingdom), 2006.

[58] S. Lecoeuche and D. Tsaptsinos. Presenting the special issue on engineering applications of neural networks - novel applications of neural networks in engineering. *Engineering Applications of Neural Networks*, 19(7):719–720, 2006.

[59] M.J. Lighthill. On sound generated aerodynamically: I general theory. In *Proceedings of the Royal Society of London A 211*, pages 564–587, 1952.

[60] D.P. Lockard and G.M. Lilley. The airframe noise reduction challenge. Technical report, NASA/TM-2004-213013, 2004.

[61] R. Lopez. Flood: An open source neural networks c++ library. www.cimne.com/flood, 2008.

[62] R. Lopez, E. Balsa-Canto, and E. Oñate. Neural networks for the solution of optimal control problems. *EUROGEN 2005 Evolutionary and Deterministic Methods for Design, Optimisation and Control with Applications to Industrial and Societal Problems*, 2005.

[63] R. Lopez, E. Balsa-Canto, and E. Oñate. Neural networks for variational problems in engineering. *International Journal for Numerical Methods in Engineering*, 75(11):1341–1360, 2008.

[64] R. Lopez, X. Diego, R. Flores, M. Chiumenti, and E. Oñate. Artificial neural networks for the solution of optimal shape design problems. In *Proceedings of the 8th World Congress on Computational Mechanics WCCM8*, 2008.

[65] R. Lopez, B. Ducoeur, M. Chiumenti, B. de Meester, and C. Agelet de Saracibar. Modeling precipitation dissolution in hardened aluminium alloys using neural networks. In *Proceedings of the 11th International Conference on Material Forming ESAFORM 2008*, 2008.

[66] R. Lopez and E. Oñate. A variational formulation for the multilayer perceptron. In *Proceedings of the 16th International Conference on Artificial Neural Networks ICANN 2006*, 2006.

[67] R. Lopez and E. Oñate. A software model for the multilayer perceptron. In *Proceedings of the International Conference on Applied Computing IADIS 2007*, 2007.

[68] R. Lopez and E. Oñate. An extended class of multilayer perceptron. *Neurocomputing*, 71(13-15):2538–2543, 2008.

[69] D.G. Luenberger. *Linear and Nonlinear Programming*. Addison Wesley, 1984.

[70] R. Luus. Application of dynamic programming to differential-algebraic process systems. *Computers and Chemical Engineering*, 17(4):373–377, 1993.

[71] D.J.C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

[72] W.S. McCulloch and W. Pitts. A logical calculus of the idea immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943.

[73] M. Minsky and S. Papert. *Perceptrons*. MIT Press, 1969.

[74] B. Mohammadi and O. Pironneau. Shape optimization in fluid mechanics. *Annual Review of Fluid Mechanics*, 36:255–279, 2004.

[75] M.F. Moller. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6:525–533, 1993.

[76] K.W. Morton and D.F. Mayers. *Numerical Solution of Partial Differential Equations, an Introduction*. Cambridge University Press, 2005.

[77] O.R. Myrh and Ø. Grong. Process modelling applied to 6082-t6 aluminium weldments - i. reaction kinetics. *Acta Metallurgica et Materialia*, 39(11):2693–2702, 1991.

[78] O.R. Myrh and Ø. Grong. Process modelling applied to 6082-t6 aluminium weldments - ii. applications of model. *Acta Metallurgica et Materialia*, 39(11):2703–2708, 1991.

[79] E. Ortega, R. Flores, and E. Oñate. An edge-based solver for compressible flow. Technical report, CIMNE, 2005.

[80] I Ortigosa, R. Lopez, and J. Garcia. A neural networks approach to residuary resistance of sailing yachts prediction. In *Proceedings of the International Conference on Marine Engineering MARINE 2007*, 2007.

[81] H. Pohlheim. Geatbx - genetic and evolutionary algorithm toolbox for use with matlab. http://www.geatbx.com, 2007.

[82] M.J.D. Powell. Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12:241–254, 1977.

[83] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press, 2002.

[84] M.M. Rai and N.K. Madavan. Aerodynamic design using neural networks. *AIAA Journal*, 38(1):173–182, 2000.

[85] A.G. Ramm. *Inverse Problems. Mathematical and Analytical Techniques with Applications to Engineering*. Springer, 2005.

[86] M. Razzaghi and S. Yousefi. Sine-cosine wavelets operational matrix of integration and its applications in the calculus of variations. *International Journal of Systems Science*, 33(10):805–810, 2002.

[87] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[88] J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language Reference Manual*. Addison Wesley, 1999.

[89] D.E. Rumelhart, G.E. Hinton, and R.J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[90] P.C. Sabatier. Past and future of inverse problems. *Journal of Mathematical Physics*, 41:4082–4124, 2000.

[91] D. Sarkar and J.M. Modak. Annsa: a hybrid artificial neural network/simulated annealing algorithm for optimal control problems. *Chemical Engineering Science*, 58:3131–3142, 2003.

[92] H.R. Shercliff, M.J. Russel, A. Taylor, and T.L. Dickerson. Microstructural modelling in friction stir welding of 2000 series aluminium alloys. *Mecanique & Industries*, 6:25–35, 2005.

[93] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, 1980.

[94] B. Stroustrup. *The C++ Programing Language*. Addison Wesley, 2000.

[95] M. Tanaka, editor. *Inverse Problems in Engineering Mechanics IV*. Elsevier, 2003.

[96] A.N. Tikhonov and V.Y. Arsenin. *Solution of ill-posed problems*. Wiley, 1977.

[97] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, 1998.

[98] J. Šíma and P. Orponen. General-purpose computation with neural networks: A survey of complexity theoretic results. *Neural Computation*, 15:2727–2778, 2003.

[99] E. W. Weisstein. Mathworld - a wolfram web resource. http://mathworld.wolfram.com, 2007.

[100] D. Whitley. The genitor algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In *Proceedings of the Third International Conference on Genetic Algorithms, San Mateo, California, USA*, pages 116–121, 1989.

[101] D.H. Wolpert and W.G. MacReady. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, 1997.

[102] O. Zienkiewicz and R. Taylor. *The Finite Element Method, Volumes 1 and 2*. Mc Graw Hill, 1988.

[103] R. Zoppoli, M. Sanguineti, and T. Parisini. Approximating networks and extended ritz method for the solution of functional optimization problems. *Journal of Optimization Theory and Applications*, 112(2):403–439, 2002.

Index