2013

Lepton Assignment Problem for the Decay of Higgs to Z Bosons to Four Leptons: A Solution Using Bayesian Neural Networks

Daniel Charles

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

LEPTON ASSIGNMENT PROBLEM FOR THE DECAY OF HIGGS TO Z BOSONS TO FOUR LEPTONS: A SOLUTION USING BAYESIAN NEURAL NETWORKS

By

DANIEL CHARLES

A Thesis submitted to the
Department of Physics
in partial fulfillment of the
requirements for the degree of
Master of Science

Degree Awarded:
Spring Semester, 2013

Daniel Charles defended this thesis on March 26, 2013.

The members of the supervisory committee were:

Harrison Prosper
Professor Directing Thesis

Todd Adams
Committee Member

Simon Capstick
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the thesis has been approved in accordance with the university requirements.

ACKNOWLEDGMENTS

I would like to thank Harrison Prosper for his guidance in my research and other academic matters, as well as his assistance in programming and using ROOT. I would also like to thank others who have assisted me, including Seth Quackenbush, Nobuo Sato, Todd Adams, Simon Capstick, Andrew Askew, Sam Bein, Arka Santra, and especially Joe Bochenek.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Within the past year, strong evidence has arisen suggesting the existence of the Higgs boson, a particle whose associated Higgs field is thought to provide mass to fundamental particles. The Higgs boson is experimentally produced at the Large Hadron Collider (LHC) by colliding together two protons. In the decay channel that we study, the protons create a Higgs boson, which then decays into two Z bosons, each of which decays into two leptons; symbolically, $pp \rightarrow H \rightarrow ZZ \rightarrow 4\ell$. The analysis that attempts to discard unrelated background data depends upon correctly assigning the leptons to their associated Z bosons. This problem is known as the lepton assignment problem. We propose a solution involving the use of a discriminant function calculated with a Bayesian Neural Network that depends on five kinematically-derived input variables.

CHAPTER 1

INTRODUCTION

Humans have speculated on the fundamental constituents of matter for thousands of years. Many ancient societies, including the Greeks, Chinese, Indians, and Egyptians, had their own theories on the matter. For example, among the ancient Greeks, the philosopher Thales believed there was only one fundamental element, water; Anaximenes believed the one element to be air; and Empedocles settled on the famous four classical elements: earth, wind, fire, and air [1].

However, these were all mere philosophical speculations, with little to no experimental reasoning to support them. Over the past few hundred years, scientists have developed theories and performed increasingly sophisticated experiments to prove the existence of the particles that make up all matter, and to determine their properties. This branch of science, as it has developed over the past century, is known as particle physics, or high-energy physics.

1.1 Particle physics: An overview of the field and the Standard Model

The idea that there exist fundamental, irreducible particles can be traced back to the ancient Greeks, as well as to later physicists and chemists, such as Galileo, Newton, Lavoisier, Dalton, and Einstein. Chemists believed at one time that each chemical element had a corresponding indivisible particle. By that reasoning, there would be well over 100 different types of fundamental particles. In the early 19th century, William Prout proposed that the hydrogen atom, being the lightest of the elements, was fundamental, and that all

other atoms were derived from it [2]. Though untrue, the idea reflected a step in the right direction.

In the late 19th century, J.J. Thomson's experiments on cathode rays demonstrated that cathode rays are charged particles, later called electrons. Thomson also measured the ratio of the electron's charge (e) to its mass (m_e). Shortly thereafter, in the early 20th century, Robert Millikan determined the value of the electron charge, e, with his famous oil drop experiment. Knowing e, as well as Thomson's ratio of e/m_e, Millikan determined the electron mass m_e, showing it to be much less than the mass of a hydrogen atom. Thus, Millikan demonstrated that there exist particles smaller than the hydrogen atom, and consequently that it must not be fundamental.

This paved the way for Ernest Rutherford's model of the atom: a small, massive, positively-charged nucleus surrounded by light, negatively-charged electrons. He called the nucleus of the hydrogen atom a proton. However, the mass of the higher elements cannot be explained solely by protons; for example, the helium nucleus has twice as much charge as hydrogen, but about four times the mass. Rutherford believed this discrepancy was due to the existence of electrically neutral particles, which he called neutrons. James Chadwick's 1932 discovery of the neutron confirmed this: atomic nuclei are indeed composed of protons and neutrons [3].

Electrons, protons, and neutrons suffice to explain the composition of atoms; the general public (and even some physicists from outside of the field) tend to be unaware of or unconcerned with the existence of any other particles. Yet the search for more fundamental particles continued. The positron was discovered in 1932; the pion and muon were discovered in 1947; and in the following decades, hundreds of other particles (many of which fell into the category of hadrons) were discovered. The resulting list of apparently fundamental particles grew so large that some came to call it "the particle zoo."

In time, physicists came to realize that not all of these particles truly were fundamental, and the list of fundamental particles grew much shorter. Almost all known particles and forces are described by the modern theory known as the Standard Model (SM). Today we claim that there are three main categories of fundamental particles: leptons (the electron, muon, tau, and their corresponding neutrinos), quarks (of which there are six types: up,

down, top, bottom, strange, and charmed), and force carriers (the W bosons, Z boson, photon, and gluon). All other particles are built from the leptons and quarks. For example, a proton consists of two up quarks and one down quark, and a neutron consists of two down quarks and one up quark. Thus, atoms fundamentally consist of quarks and electrons. Additionally, all of the known forces (the electromagnetic, strong, and weak forces, and gravity) are mediated by a corresponding force carrier particle [4, 5], though this has yet to be confirmed for gravity.

In addition to describing all of the fundamental particles and forces, the SM combines the theory of the electromagnetic force with that of the weak force. However, a mathematical consequence of this unification is that the weak force's associated particles (the W and Z bosons) have no mass. Experiments, on the other hand, have shown that this is not the case: the masses of the W and Z bosons are 81.4 GeV/c^2 and 91.2 GeV/c^2, respectively [6]. For comparison, the mass of the electron is 0.511 MeV/c^2; relative to the electron, the W and Z bosons are quite massive.

1.2 The Higgs boson

To correct this theoretical problem, physicists proposed the existence of a field that is present throughout all of space, known as the Higgs field. The mass of an elementary particle is due to its interaction with the Higgs field, and particles interact with the Higgs field via its corresponding particle, the Higgs boson [7]. This particle has proven difficult to observe. Convincing evidence for the existence of a particle consistent with the Higgs boson was announced in 2012, but experiments to determine its properties remain ongoing.

1.2.1 Experimental difficulties in the Higgs search

Physicists were able to observe the proton and electron long ago, by means of relatively simple experiments. One crucial reason that this was possible is that the proton and electron are stable particles. Much like chemical atoms, some particles are stable, while others decay rapidly into other particles. For example, protons and electrons have never been observed to decay, whereas neutrons outside of the nucleus decay with a mean lifetime of approximately 881 seconds, or about fifteen minutes [8].

Many fundamental particles, including the Higgs boson, have much shorter lifetimes. A particle's mean lifetime can be calculated by determining the distribution of its mass. The result of many measurements of such a particle's mass is a Breit-Wigner distribution, which is similar in shape to a Gaussian, with a width at half-maximum of Γ, as in Fig. 1.1. From this, a particle's mean lifetime τ is given by $\tau = \hbar/\Gamma$ [9]. Using this equation, as well as the measured width of the Z boson, $\Gamma_Z = 2.495$ GeV/c^2 [6], we obtain $\tau_Z = 2.638 \times 10^{-25}$ s, for the Z boson. For the Higgs boson, assuming a mass of $m_H = 126$ GeV/c^2, its width has been theoretically calculated [10] to be 4.21×10^{-3} GeV/c^2, and so $\tau_H = 1.56 \times 10^{-22}$ s.

Due to its very short lifetime, the Higgs boson cannot be observed directly. We can, however, infer its existence through its decays into other particles. In practice, we can create a Higgs boson by colliding two protons at high energies, such as at the Large Hadron Collider (LHC). Once a Higgs is created, there are numerous different mechanisms by which it can decay. Two of these decays are the most relevant to the Higgs search. In the first type, the Higgs boson decays into two photons, or, symbolically, $H \to \gamma\gamma$. In the second type, which is the topic of this thesis, the Higgs boson decays into two Z bosons, each of which decays into two electrons or two muons: $H \to ZZ \to 4\ell$ [11], where ℓ represents a lepton (see Fig. 1.2).

Our approach to searching for the Higgs is to search for events in which we observe four leptons, in the hopes that those leptons arose from a Higgs decay. However, other types of decays can also produce four leptons. For example, a collision of two protons can give rise to two Z bosons, each of which decays into two leptons: $pp \to ZZ \to 4\ell$. In both this decay mechanism (the "background") and the mechanism that we are searching for (the "signal"), the end result that we observe is four leptons. Our goal, then, is to find some way to identify and discard background events. In doing so, we increase the statistical significance of our observation of the signal.

1.2.2 Prior experiments

The theory that describes the interactions of the Higgs boson does not give a value for its mass. The mass of the Higgs boson, m_H, is a free parameter. Determining this value has been a top priority in the Higgs search.

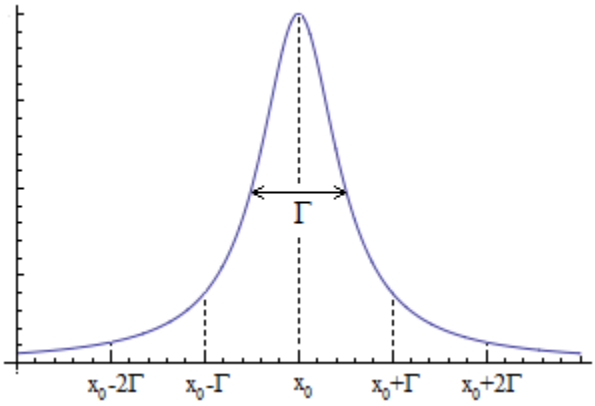


Figure 1.1: Breit-Wigner distribution showing mass distribution of fundamental particles, centered about x_0.

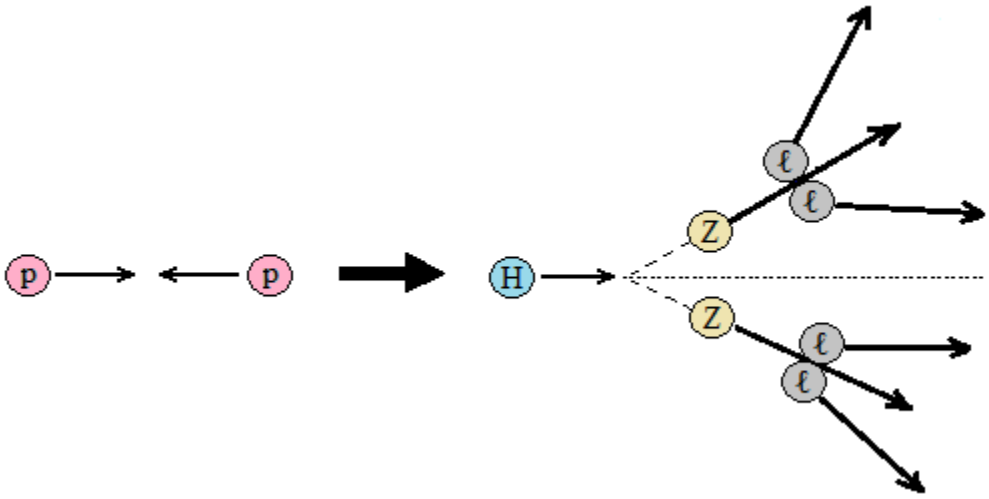


Figure 1.2: The signal decay process. Two protons collide (left), decaying to a Higgs, then into Z bosons, then into leptons.

The search for the Higgs boson was not performed with a single experiment; physicists have searched for it for decades. Over time, experiments have added more and more restrictions on the possible value of its mass. For example, experiments at the Large Electron Positron Collider (LEP) placed a lower bound of 114.4 GeV/c^2 on the Higgs boson mass [12]. Analysis of other experiments at the Stanford Linear Accelerator Center (SLAC) and LEP assigned an upper bound of 158 GeV/c^2 [13].

Many of the initial searches were performed at LEP (the predecessor to the LHC) and at the Tevatron. LEP, located at CERN (the European Organization for Nuclear Research) in Switzerland, was constructed in the 1980s. It was a collider based on a circular accelerator ring, 27 km in circumference, that collided electrons and positrons. Its experiments ran from 1989 to 2000, after which it was disassembled for the LHC to take its place. Its maximum center-of-mass energy was 209 GeV/c^2 [14]. Experiments at LEP hinted at a possible signal of the Higgs boson [15], with a mass $m_H = 115.6$ GeV/c^2. However, this result was not statistically significant enough to announce a discovery, and the range of 110 to 121.5 GeV/c^2 has since been excluded [11]. The Tevatron, located at Fermilab in Illinois, was also a circular accelerator. Its experiments ran from 1985 to 2011 [16], with a maximum center-of-mass energy of 1.96 TeV [17].

At present, experiments in the search for the Higgs boson are performed at the LHC. The LHC uses the same tunnel previously occupied by LEP. Unlike LEP, which collided electrons and positrons, the LHC collides protons with protons. (In other experiments unrelated to the Higgs search, the LHC also collides heavy ions, such as lead.) The beams of protons travel independently around the 27 km circular path, intersecting only at four fixed locations. The main detectors relevant to the Higgs search, the Compact Muon Solenoid (CMS) and A Toroidal LHC Apparatus (ATLAS), are placed at two opposing intersection points [18].

On July 4, 2012, scientists at CERN announced to the public the discovery of "a new particle, at the [significance] level of 5 sigma, in the mass region around 126 GeV" [19]. The CMS and ATLAS Collaborations published journal articles that August discussing their results [11, 20]. Although the scientists have not yet officially declared this particle to be

the Higgs boson of the SM, we have proceeded with our work as though it is indeed the SM Higgs, with a mass of 125 GeV/c^2.

1.2.3 The lepton assignment problem

This thesis will focus on the decay channel $H \to ZZ \to 4\ell$. Recall that the only particles in this decay that the LHC detectors observe directly are the leptons. The current analysis that attempts to eliminate background events requires that we correctly pair the leptons to their parent Z bosons. This is not a trivial task.

Recall that leptons are electrons (e^-), muons (μ^-), taus (τ^-), and their respective antiparticles (e^+, μ^+, and τ^+). In the SM, when a Z boson decays into two leptons, the leptons must be of the same flavor and opposite charge. For example, $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ are acceptable, whereas $Z \to e^+\mu^-$ (different flavor) and $Z \to e^+e^+$ (same charge) are not.

Given these rules, we can classify the decay of the Z bosons into two categories: a decay into two pairs of leptons of different flavors, or a decay into two pairs that are all of the same flavor. As an example of the former, if we observe $ZZ \to e^+e^-\mu^+\mu^-$, then we can say outright that the electron and positron must have decayed from one of the Z bosons, and the muons must have decayed from the other.

However, if we observe $ZZ \to \mu^+\mu^-\mu^+\mu^-$, we cannot say *a priori* which of the muons should be paired together. Generally, given four objects, there are three possible ways to pair them into groups of two, as in Fig. 1.3. However, given the restriction that a Z boson can only decay into two particles of opposite charge, there remain only two possible pairings, as in Fig. 1.4.

One possible pairing is that the first and second muon decayed from one Z boson, and the third and fourth muon decayed from the other Z boson. Alternatively, it is just as possible that the first and fourth muon decayed from one Z, and the second and third muon decayed from the other. Further analysis is necessary to resolve this ambiguity, and determine which leptons came from which Z boson. We refer to this problem as the lepton assignment problem; it is the main topic of this thesis. Our objective is to find some function of the leptons' four-vectors that can help us decide which of these two possible pairings is the correct one.

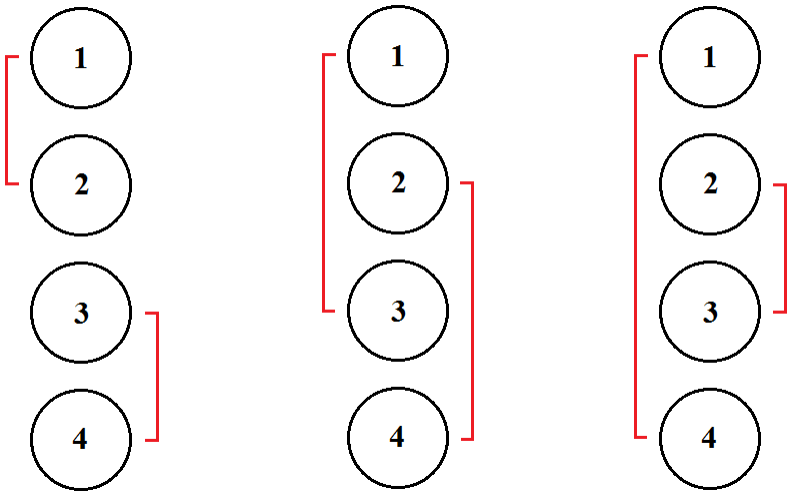
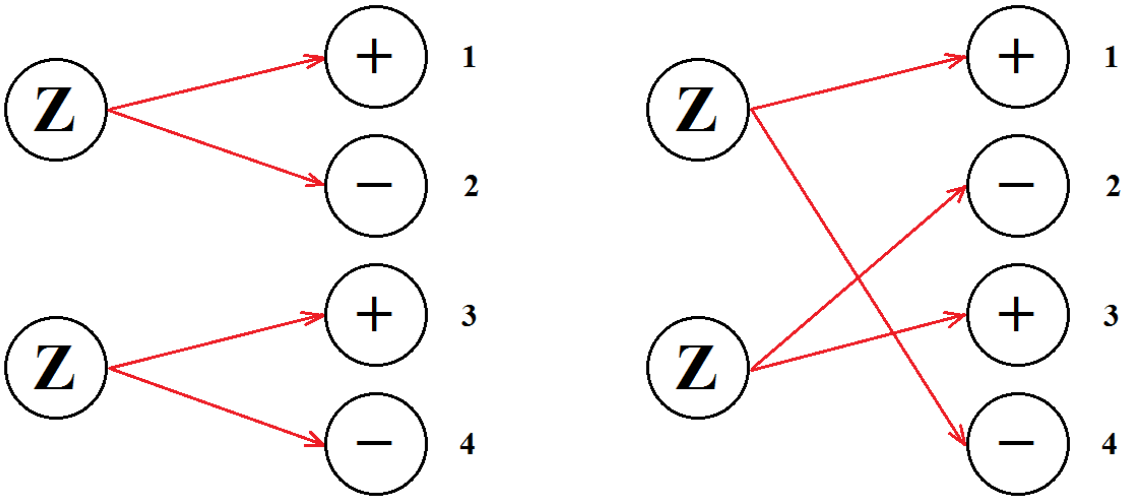


Figure 1.3: Three possible pairings of four objects.



(a) (b)

Figure 1.4: Possible decays of two Z bosons into four leptons of the same flavor.

1.3 Conventions and units

In high energy physics, energies are commonly given in units of electron-volts (eV), rather than SI units of joules (J). To convert these energies into SI units, recall the relation $1\text{ eV} = 1.602 \times 10^{-19}$ J. These units are used because of the low energies present in collisions, relative to 1 J. Despite the relativistic speeds of the particles involved, even very high energy collisions have beam energies on the order of TeV, which is small relative to 1 J. For example, $1\text{ TeV} = 1.602 \times 10^{-7}$ J.

Additionally, it is conventional to state equations and quantities in units that do not contain the constants $\hbar$ or c. For example, the relativistic equation relating energy (E), momentum (p), and mass (m), $E^2 = p^2c^2 + (mc^2)^2$, would be written in these units as $E^2 = p^2 + m^2$. In practice, this means that energy, momentum, and mass all have the same units, eV. When returning to standard units, recall that the units of momentum are eV/c, and the units of mass are eV/c^2.

Note also that a particle's momentum is not typically described in terms of its momentum vector $\vec{p}$. Instead, we describe the particle by its momentum transverse to the plane of the collision (p_T), the angle made by $\vec{p}_T$ with respect to the vertical (φ), and its pseudorapidity (η); see Fig. 1.5. Pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, where θ is the angle made by $\vec{p}$ with respect to the beam axis [21]. When $\theta \to 0$, $\eta \to \infty$; and when $\theta = \pi/2$, $\eta = 0$. Thus, as θ goes from 0 to $\pi/2$, η goes from $+\infty$ to 0.

If we wish to relate these three variables (p_T, η, and φ) to the Cartesian components of the momentum vector (p_x, p_y, and p_z), the following relations [21] can be used:

$$p_x = p_T \cos\varphi$$

$$p_y = p_T \sin\varphi$$

$$p_z = p_T \sinh\eta$$

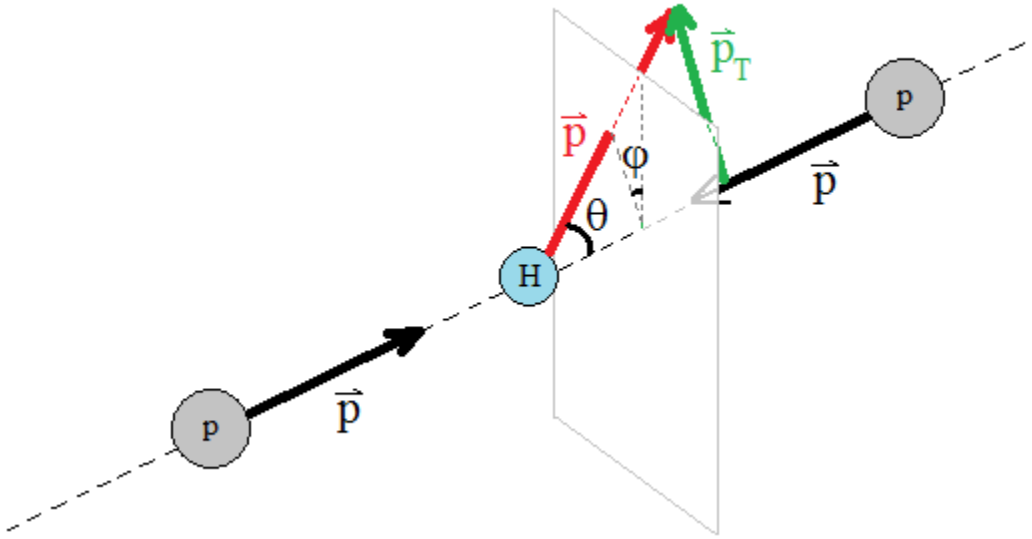


Figure 1.5: Variables used in particle physics: $|\vec{p}_T|$, θ, and φ.

CHAPTER 2

A SIMPLE EXAMPLE: DISCRIMINANT USING ONLY Z BOSON MASSES AS INPUT

In order to solve the lepton assignment problem, we will make use of a so-called *discriminant function*, $D(\vec{x})$. By feeding this function a set of input values, $\vec{x}$, we can use the output value $D(\vec{x})$ to classify the input into one of two categories: the correct pairing, or the incorrect pairing. The act of classifying into two categories is known as binary classification [22].

Our solution to the lepton assignment problem will make use of a discriminant function that depends on a set of five input variables, $\vec{x}$. Before describing the full solution, let us first discuss a more straightforward example: classifying events as either signal or background, using a discriminant depending on only two variables, the Z boson masses.

2.1 Technical details

2.1.1 Data files

Once we create an algorithm to perform the lepton assignment, that algorithm will be applied to real data obtained from experiments, such as those at the LHC. At this stage, however, this analysis was performed using simulated data.

The data can be split into two categories: *generator-level* and *reconstruction-level* data. The reconstruction level simulates collision events, as well as how a detector will respond to those events. The generator level simulates collisions on a purely theoretical level, without accounting for the detector's response; this is the type of data that we will be using.

One advantage to this approach is that we can ignore any complications that might arise from a detector's response. Additionally, the generator-level data tell us all of the details

for an event. Because detectors observe only the leptons directly, the reconstruction-level data contain no explicit information about the Higgs or Z bosons. The generator-level data provide parameters (including p_T, η, φ, energy, and charge) for all of the particles: the Higgs boson, the Z bosons that it decays into, and the leptons that they decay into.

2.1.2 ROOT

All of the programming in this thesis was performed using Python. The ROOT system was also used extensively within Python. ROOT can be used to create data files, open them, create histograms, and perform computations. The tree-based file system used by ROOT allows for efficient access of data [23]. This is extremely important in experiments at the LHC, in which vast amounts of data are stored and accessed. ROOT also contains classes for vectors with two or three components, as well as Lorentz four-vectors. By using these vector classes, vector calculations such as the dot product or finding the invariant mass become much simpler.

All of the data files that we used are ROOT files. There are separate files for both the signal ($pp \to H \to ZZ \to 4\ell$) and the background ($pp \to ZZ \to 4\ell$). Within each of these files, there are variables for both generator- and reconstruction-level data. There are also several different types of generator-level values available in these files. These different values mostly contain the same information, just ordered in a different way. These variables are of the form MC__*varname*, MC__LEPT__*varname*, and MC__ZZ__*varname*. For example, MC__ZZ__PT is a variable for p_T, and MC__ZZ__ETA is for η.

Due to the tree-based structure of ROOT files, a variable MC__ZZ__*varname* essentially behaves like a list. The value of that parameter for a certain particle can be obtained by specifying an index. For example, MC__ZZ__PT[i] gives the value of p_T for the particle at index i, where i is an integer. A branch MC__ZZ__*varname* is structured such that $i = 0$ represents the Higgs boson, $i = 1$ and 2 represent the Z bosons, $i = 3$ and 4 represent the leptons that decayed from the Z boson in $i = 1$, and $i = 5$ and 6 represent the leptons that decayed from the Z boson in $i = 2$. We chose to use these MC__ZZ__*varname* variables because this ordering is particularly convenient for the lepton assignment problem.

2.2 Calculating the discriminant explicitly

We will later use a discriminant function to determine the correct pairings of leptons. However, a more common use of the discriminant is to classify an event as belonging to either the signal or the background. The range of possible output values for $D(\vec{x})$ is [0, 1]. A value near 1 indicates that the event most likely came from the signal, and a value near 0 indicates that the event most likely came from the background. Graphically, the discriminant should be similar in form to Fig. 2.1.

The discriminant function can be written explicitly as follows:

$$\begin{aligned} D(\vec{x}) &= \frac{f(\vec{x}|S)}{f(\vec{x}|S) + f(\vec{x}|B)} \\ &= \frac{S(\vec{x})}{S(\vec{x}) + B(\vec{x})}. \end{aligned} \tag{2.1}$$

The notation $f(A|B)$ represents a probability density of "A, given B." $S(\vec{x}) = f(\vec{x}|S)$ represents the probability density that an event is described by the variables $\vec{x}$, given that the event comes from the signal. $B(\vec{x}) = f(\vec{x}|B)$ represents the probability density that an event is described by the variables $\vec{x}$, given that the event comes from the background [24].

Consider a discriminant function $D(\vec{x})$ that depends only on two variables, $\vec{x} = (x_1, x_2)$. If we have a large sample of values for x_1 and x_2 for both the signal and the background, we can plot those values as two-dimensional histograms, showing the distributions of x_2 versus x_1. These histograms would then represent $S(\vec{x})$ and $B(\vec{x})$. By performing algebra on these histograms, we can explicitly calculate $D(\vec{x})$ via Eq. 2.1.

This was exactly the technique that we used as a first step, to create a simple discriminant function. We extracted the invariant masses for the two Z bosons from our data files, for both the signal and the background. We then labeled the two Z bosons such that the one with the larger mass was Z_1, and the one with the smaller mass was Z_2. The resulting two-dimensional histograms for m_{Z_2} versus m_{Z_1} are shown in Fig. 2.2.

Note that the signal has a peak near (m_{Z_1}, m_{Z_2}) = (90 GeV, 25 GeV). Since the measured Z mass is around 90 GeV, one might have expected this peak to be near (90 GeV, 90 GeV). However, because the two Z bosons decayed from a Higgs boson with a mass of 125 GeV, conservation of energy implies that the two Z bosons cannot both have a mass of

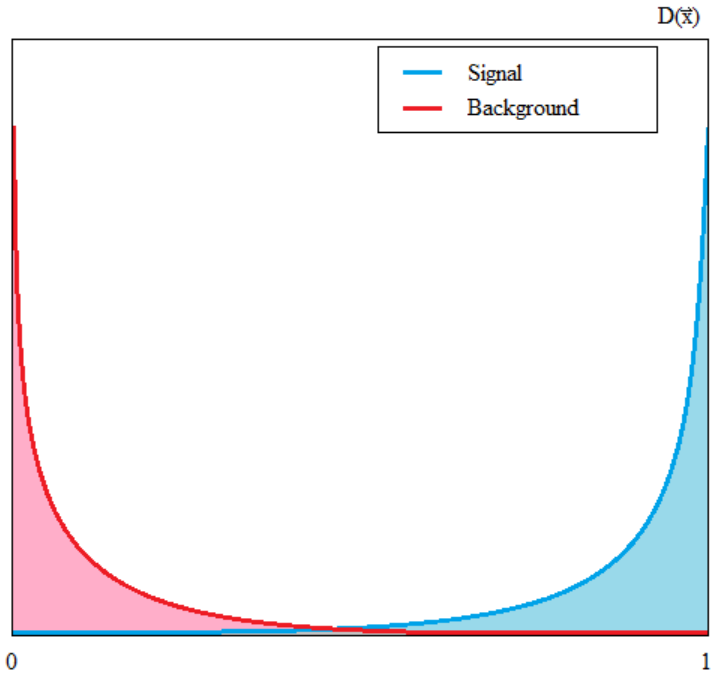
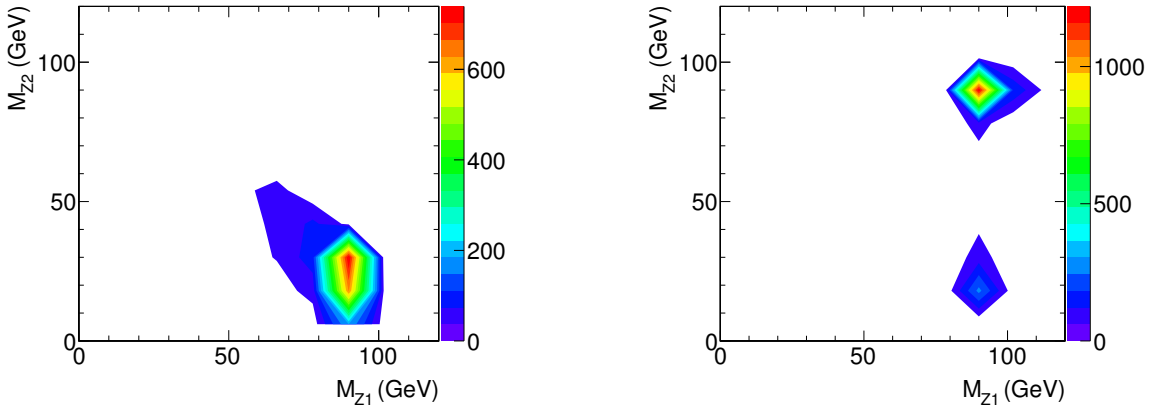


Figure 2.1: Histogram showing the desired form of a discriminant function, $D(\vec{x})$.



(a) Signal: $pp \to H \to ZZ \to 4\ell$

(b) Background: $pp \to ZZ \to 4\ell$

Figure 2.2: Histograms of m_{Z_2} vs. m_{Z_1} for signal and background. The values on the z axis represent the number of events.

90 GeV; otherwise, the total energy would exceed 125 GeV. Thus, the Higgs mass of 125 GeV places a restriction on the Z masses. Statistically, the most probable outcome is that one Z boson (which is said to be *on-shell*) will have a mass near 90 GeV, and the remaining energy (and thus mass) will go to the other (*off-shell*) Z boson.

This differs from the background, which has a peak value near (90 GeV, 90 GeV). This makes sense because each of the Z bosons are created independently, directly from the collision of the protons. Because the Z bosons do not decay from another particle, as they do in the signal, we do not see that restriction on the two masses.

Now that we have the distributions for $S(\vec{x})$ and $B(\vec{x})$, in the form of histograms, we can calculate the discriminant explicitly using Eq. 2.1. The resulting histogram is shown in Fig. 2.3a.

2.3 Bayesian neural networks

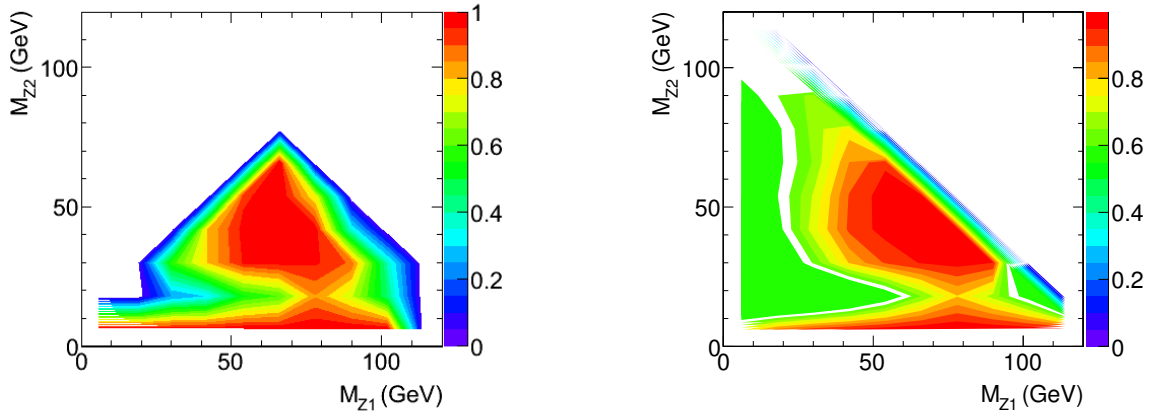
There are limitations to calculating the discriminant as we did above. The use of finite data sets, as well as the presence of random noise within the data, will often result in a noisy discriminant. Instead of calculating the distributions $S(\vec{x})$ and $B(\vec{x})$ and then calculating $D(\vec{x})$ directly, it is preferable to attempt to model their behavior by fitting the discriminant with a smooth function.

Unfortunately, although a linear one-dimensional fit is simple to perform, it is much more difficult to provide a fit for a more complicated function. Additionally, we often use more than one input variable, raising the difficulty further, as we are now attempting to fit a complicated function in many dimensions.

There are many different computational techniques by which we can approximate a discriminant function. Among them are boosted decision trees (BDT), multilayer perceptrons (MLP), neural networks (NN), and Bayesian neural networks (BNN). All of these methods attempt to approximate the same function, $D(\vec{x})$; they simply use different techniques to do so. We will be using the last method, the BNN.

A Bayesian neural network is similar to a neural network [22],

$$n(\vec{x}, \vec{\omega}) = \frac{1}{1 + e^{-f(\vec{x}, \vec{\omega})}}, \tag{2.2}$$



(a) $D(\vec{x})$ calculated explicitly with Eq. 2.1

(b) $D(\vec{x})$ computed with BNN software

Figure 2.3: Histograms of the discriminant function $D(m_{Z_1}, m_{Z_2})$. The values on the z axis represent the value of the discriminant function.

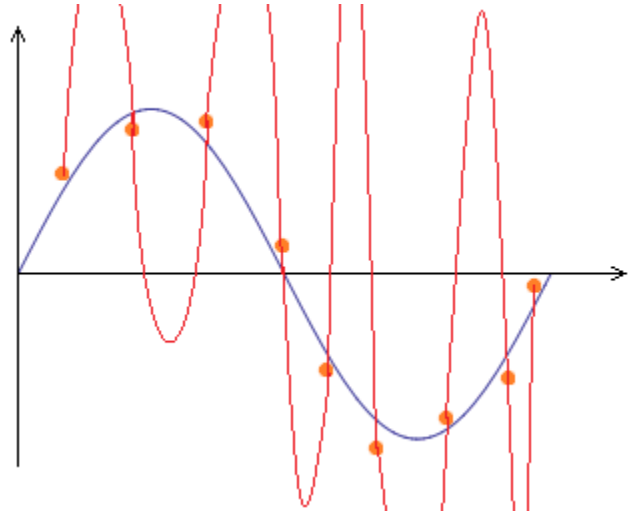


Figure 2.4: Demonstration of overtraining for a fitted function. The orange dots are the data, the blue line is the underlying function used to generate that data, and the red line is an overtrained function.

where

$$f(\vec{x},\vec{\omega}) = a + \sum_{j=1}^{H} b_j \tanh(c_j + \sum_{i=1}^{P} d_{ji} x_i). \tag{2.3}$$

H is the number of *hidden nodes* and P is the number of input parameters. A neural network $n(\vec{x},\vec{\omega})$ can be used to model a discriminant function. In fact, a neural network can model any real function, provided the number of hidden nodes H is large enough. The computational challenge lies in calculating the function $f(\vec{x},\vec{\omega})$. This function depends on both our input values, $\vec{x}$, and on some other arbitrary parameters, $\vec{\omega} = (a, b_j, c_j, d_{ji})$, that we vary.

An ideal discriminant function $D(\vec{x}) = n(\vec{x},\vec{\omega})$ would yield a value close to 1 for a signal event, and close to 0 for a background event. However, for some values of our inputs $\vec{x}$, this function $D(\vec{x})$ will not behave as desired; for example, perhaps a particular signal event will give $D(\vec{x})$ near 0. Such problems are unavoidable, as our classification techniques are not perfect. Also, it is possible for certain signal and background events to both be described by the same parameters; thus, some events simply cannot be distinguished at all. Although misclassifications cannot be eliminated completely, we would like to minimize how often they occur.

In solving a classification problem, it is common to define an error function $E(\vec{\omega})$, which quantifies how much the input data deviate from the fitted function. In constructing a neural network, one attempts to fit a function $n(\vec{x},\vec{\omega})$ to the data by adjusting the values of the parameters $\vec{\omega}$. If the software can find the values $\vec{\omega} = \vec{\omega}_0$ that minimize the error function $E(\vec{\omega})$, the problem is solved: plug these values $\vec{\omega}_0$ into $f(\vec{x},\vec{\omega})$ (Eq. 2.3), and use Eq. 2.2 to calculate $n(\vec{x},\vec{\omega}_0)$. The result $n(\vec{x},\vec{\omega}_0)$ is a neural network with the least possible classification error.

A potential concern with fitting functions to data is that one might *overtrain* the function. That is, one might create a function that perfectly fits all of the data points, but does not capture the underlying trend. For example, suppose we have a set of data points that represent a sine function, with some noise. Despite the noise in the data, the overall functional form is sinusoidal; a fitted function should approximate that general trend (see Fig. 2.4).

However, it is possible to tune the coefficients of a high-order polynomial such that the polynomial passes through all of the data points. Such a polynomial would give an error function $E(\vec{\omega})$ of zero, but this function clearly does not represent the overall trend. Similar issues can also arise when using neural networks. We attempt to limit this problem by dividing our data samples into two smaller samples: a *training sample* and a *testing sample*. The neural network software attempts to fit a function to the training sample, and then evaluates the accuracy of that function on the testing sample. If the fitted function performs poorly on the testing sample, overtraining is a likely cause.

Bayesian neural networks (BNNs) are effective at reducing overtraining. A BNN is essentially an average of neural networks. The relationship between a BNN $\overline{n(\vec{x})}$ and a neural network $n(\vec{x}, \vec{\omega})$ is given [24] by:

$$\begin{aligned} \overline{n(\vec{x})} &= \int n(\vec{x}, \vec{\omega})\, p(\vec{\omega}|t)\, d\vec{\omega} \\ &\approx \frac{1}{K} \sum_{k=1}^{K} n(\vec{x}, \vec{w_k}) \end{aligned} \tag{2.4}$$

Analytically, a BNN is a weighted integral of neural networks, weighted by a probability density function $p(\vec{\omega}|t)$. However, such an integral cannot be evaluated directly. Instead, the BNN software approximates this integral numerically, using the Hybrid Markov Chain Monte Carlo algorithm [25] to select K representative values of $\vec{\omega}$. Each set of values $\vec{\omega}_k$ corresponds to a particular neural network; the BNN is calculated by summing over these individual neural networks, as in Eq. 2.4.

2.4 Creating the discriminant with BNN software

We will now create a discriminant function $D(m_{Z_1}, m_{Z_2})$ using BNN software adapted from Radford Neal's Flexible Bayesian Modeling (FBM) package. If successful, the result should appear very similar to the explicitly calculated discriminant in the previous section (Fig. 2.3a).

To create the discriminant function, we first read out the values of m_{Z_1} and m_{Z_2} for the signal and background, remembering to order them as before (such that Z_1 has the larger mass). We create text files for both the signal and background, with columns for m_{Z_1} and

m_{Z_2}, and another column containing either a 1 to indicate that the event is from the signal, or a 0 indicating that the event is from the background. After combining and mixing the text files, we use the resulting data file as input for the BNN software. Once the software finishes running, the final output is a C++ file that encodes our discriminant function. If we call this C++ function with values of m_{Z_1} and m_{Z_2} as arguments, the output $D(m_{Z_1}, m_{Z_2})$ will be some number between 0 and 1.

Upon obtaining the discriminant function, we plotted a two-dimensional histogram of it. This was done by setting up an empty histogram, and discretizing the x and y axes into a grid. The value of the histogram for a given bin was taken to be the value of the discriminant function at the center of that bin. We looped over all values of x and y (that is, m_{Z_1} and m_{Z_2}), calculating $D(\vec{x})$ for each bin with our C++ function; Figure 2.3b shows the resulting plot.

We see that both discriminant functions are of roughly the same form. Both have peaks around (60 GeV, 45 GeV) which drop off in all directions. In practice, one could use such a plot to weight the likelihood that a given event came from the signal. For example, if the event lies in a colorless region (say, in the top-right corner), then $D(\vec{x}) \approx 0$, and the event was most likely from the background. If, however, the event lies in a red region (say, near (60 GeV, 45 GeV)), then $D(\vec{x}) \approx 1$, and the event was most likely from the signal. This result is a simple method for eliminating background events.

CHAPTER 3

SOLUTION OF THE LEPTON ASSIGNMENT PROBLEM

The previous chapter discussed the use of a BNN to distinguish between signal and background events. This chapter discusses our algorithm for solving the lepton assignment problem, by means of a BNN discriminant function.

Our method for creating this lepton assignment algorithm was similar to the method for creating the previous BNN that was used to distinguish signal and background. We attempted to find a set of input variables $\vec{x}$ that had significantly different distributions when calculated for the correct pairing and for the incorrect pairing. Many different variables were examined; in the end, we selected five of these variables as inputs for our BNN discriminant.

3.1 Choosing the input variables

Recall that our data files contain the variables p_T, η, and φ for all four leptons. From these variables, we can obtain the Cartesian momentum vector $\vec{p} = p_x\,\hat{x} + p_y\,\hat{y} + p_z\,\hat{z}$, or, equivalently, $\vec{p} = \vec{p}_T + p_z\,\hat{z}$, which we will use to calculate most of our input variables.

Also recall that in our ROOT files, the leptons are ordered such that the leptons in position $i = 3$ and 4 correspond to the Z boson in $i = 1$, and the leptons in $i = 5$ and 6 correspond to the Z boson in $i = 2$. For convenience, we will refer to the first two leptons as #1 and #2, and the second two as #3 and #4. Additionally, under this labeling scheme, we number them such that the first lepton in each pair (#1 and #3) is the positive one, and the second lepton (#2 and #4) is the negative one.

3.1.1 Cosine of angle between momentum vectors

The first variable that we examined is $\cos\theta_{ij}$, the cosine of the angle between the momentum vectors $\vec{p}$ of the leptons. This was found using the relation $\vec{p_i} \cdot \vec{p_j} = |\vec{p_i}||\vec{p_j}| \cos\theta_{ij}$:

$$\cos\theta_{ij} = \frac{\vec{p_i} \cdot \vec{p_j}}{|\vec{p_i}||\vec{p_j}|}. \tag{3.1}$$

Using this equation, we calculated $\cos\theta_{ij}$ for the correct pairing: for leptons #1 and #2 ($\cos\theta_{12}$), and for leptons #3 and #4 ($\cos\theta_{34}$). Then, we performed the same calculation, except deliberately using the incorrect lepton pairing: leptons #1 and #4 ($\cos\theta_{14}$), and #2 and #3 ($\cos\theta_{23}$). It is important to use these particular pairings when performing calculations for the incorrect pairing, to maintain conservation of charge. That is, it would be invalid to pair together #1 and #3, as both are positively charged. The same applies for #2 and #4, which are both negative.

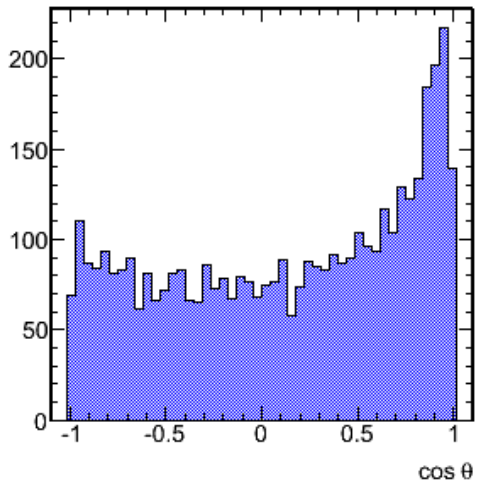
We then filled a histogram of the variable $\cos\theta$ for both the correct and incorrect pairings. For the correct pairing, the values of $\cos\theta_{12}$ and $\cos\theta_{34}$ were filled in the same histogram. For the incorrect pairing, the values of $\cos\theta_{14}$ and $\cos\theta_{23}$ were filled. These histograms are shown in Fig. 3.1. We see that these distributions are somewhat different, but not as distinctly as we would like.

3.1.2 Cosine of the angle between transverse momentum vectors

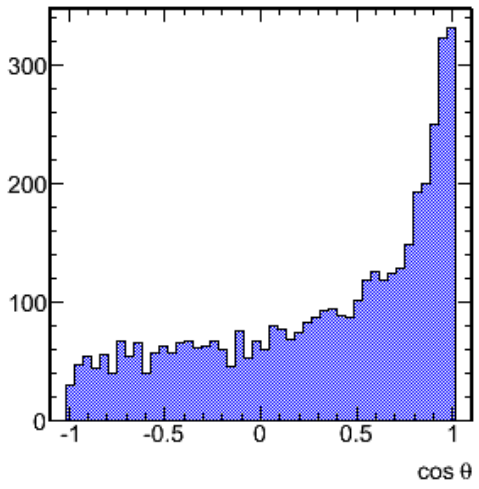
Next we examined the cosine of the angle between the leptons' transverse momentum vectors $\vec{p}_T$. The same formula as before was used (Eq. 3.1), except we replaced the vectors $\vec{p}$ with the transverse vectors $\vec{p}_T$ (recalling that $\vec{p}_T = p_x\,\hat{x} + p_y\,\hat{y}$). The histograms for this variable $\cos\theta_T$ are shown in Fig. 3.2. These distributions provide a much clearer distinction between the correct and incorrect pairings than did the distributions for $\cos\theta$. Thus we selected this variable, $\cos\theta_T$, as the first of our input variables $\vec{x}$.

3.1.3 Difference between transverse momenta

Our next input variable is Δp_T, the difference between the transverse momenta of a given lepton pair. For the correct pairings, we used $\Delta p_{T_{12}} = |p_{T_1} - p_{T_2}|$ and $\Delta p_{T_{34}} = |p_{T_3} - p_{T_4}|$. We took the absolute value so that all of these values are positive. Similarly, for the incorrect pairings, we calculated $\Delta p_{T_{14}}$ and $\Delta p_{T_{23}}$.

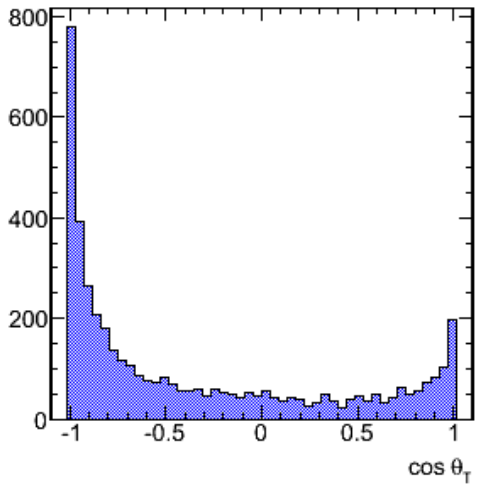


(a) Correct pairing

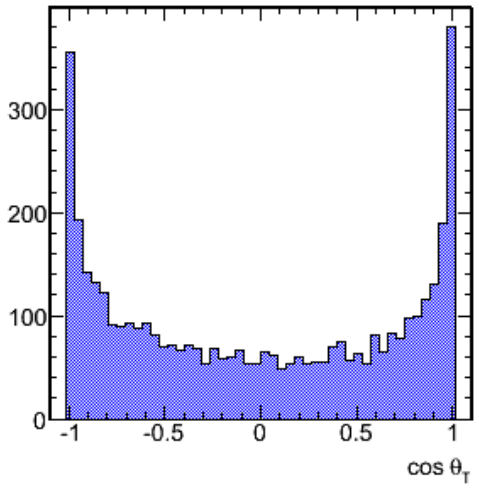


(b) Incorrect pairing

Figure 3.1: Histograms of $\cos\theta$ for correct and incorrect pairings



(a) Correct pairing



(b) Incorrect pairing

Figure 3.2: Histograms of $\cos\theta_T$ for correct and incorrect pairings

We filled the histogram for the correct pairing with $\Delta p_{T_{12}}$ and $\Delta p_{T_{34}}$, and for the incorrect pairing with $\Delta p_{T_{14}}$ and $\Delta p_{T_{23}}$; see Fig. 3.3. Again, this provides a clear distinction between correct and incorrect pairings. The distribution for the correct pairing drops off immediately from $\Delta p_T = 0$ GeV, whereas the distribution for the incorrect pairing peaks near $\Delta p_T = 25$ GeV.

3.1.4 Invariant mass

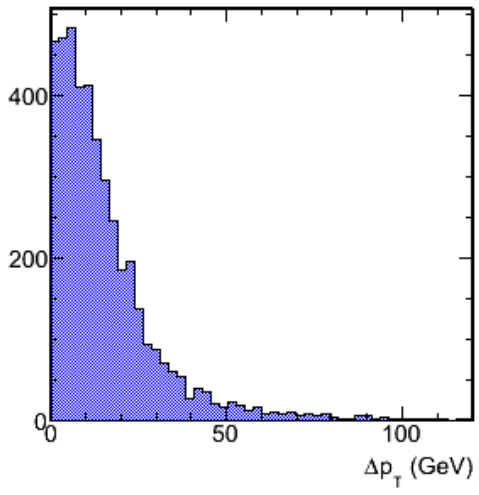
The next variable that we examined was the invariant mass of a particular pair. In our units where $c = 1$, the invariant mass of a particle is equal to the magnitude of its momentum four-vector. To find the invariant mass of multiple particles, one adds their momentum four-vectors, and computes the magnitude of the resultant, $\sqrt{p^\mu p_\mu} = \sqrt{E^2 - \vec{p} \cdot \vec{p}}$. In this particular problem, the invariant mass of a correct pairing (e.g., leptons #1 and #2) equals the invariant mass of their corresponding Z boson. Such a calculation for an incorrect pairing (e.g., leptons #1 and #4) would be physically meaningless, and the distribution of the resulting values would likely be quite different.

We calculated the invariant mass for the correct pairs, m_{12} and m_{34}, and for the incorrect pairs, m_{14} and m_{23}; the resulting histograms are shown in Fig. 3.4. For the correct pairing, we see in Fig. 3.4a that there is a sharp peak near 90 GeV (corresponding to the on-shell Z boson), and a broader peak around 25 GeV (the off-shell Z boson). Note that this behavior agrees with that of our previous plot in Fig. 2.2a. For the incorrect pairing, the plot for the invariant mass (Fig. 3.4b) makes no physical sense; there is only a single broad peak near 40 GeV. Though these distributions are quite different, they also have quite a bit of overlap. We did not choose to use this as one of our input variables.

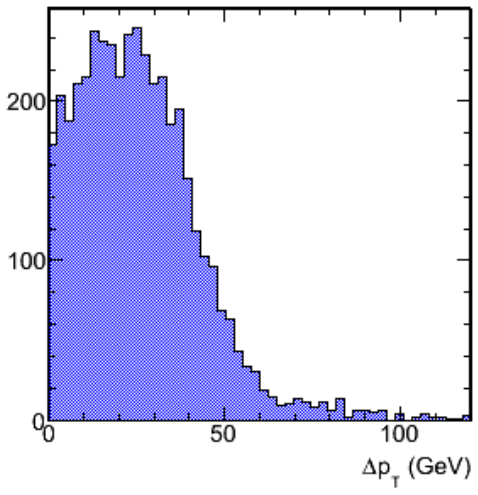
3.1.5 Ratio of invariant masses

Although we did not use the invariant mass as an input variable, we used it to calculate a variable that we did use. Recall that the previous calculation yielded two invariant masses for each pairing: m_{12} and m_{34} for the correct pairing, and m_{14} and m_{23} for the incorrect pairing. Our next input variable is simply the ratio of the two invariant masses.

In order to obtain a ratio with a value between 0 and 1, let the smaller of the two masses be M_1 and the larger be M_2; the ratio that we calculate is M_1/M_2. For example, suppose

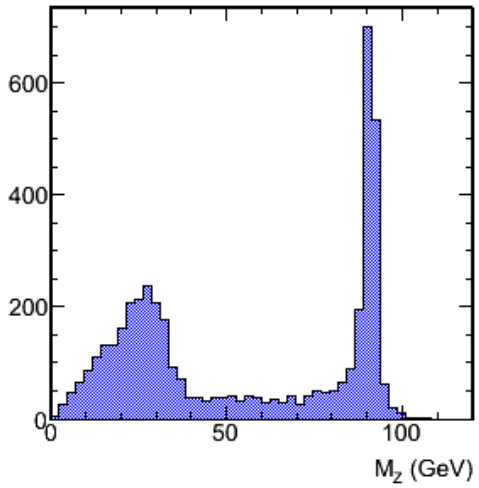


(a) Correct pairing

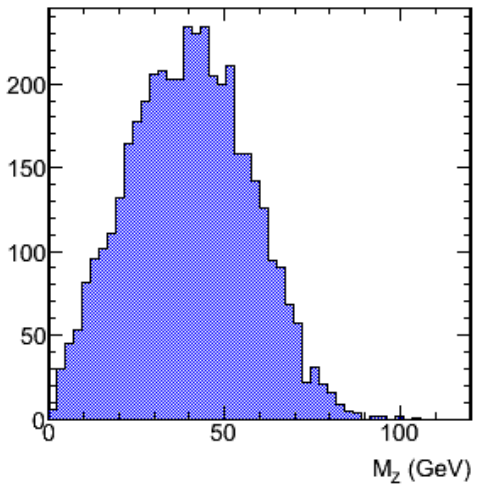


(b) Incorrect pairing

Figure 3.3: Histograms of Δp_T for correct and incorrect pairings



(a) Correct pairing



(b) Incorrect pairing

Figure 3.4: Histograms of invariant masses m_{ij} for correct and incorrect pairings

we find for the correct pairing that $m_{12} > m_{34}$. We label the smaller mass m_{34} as M_1 and the larger mass m_{12} as M_2, and so our variable, the ratio M_1/M_2, is m_{34}/m_{12}.

The histograms for this ratio are shown in Fig. 3.5. The plot for the correct pairing has a peak near 0.3, and the plot for the incorrect pairing has a broader peak around 0.7. Although there is some overlap between them, the histograms of the ratio provide a clearer distinction than those of the masses alone. Thus we selected this ratio as the last of the input variables $\vec{x}$ for our BNN.

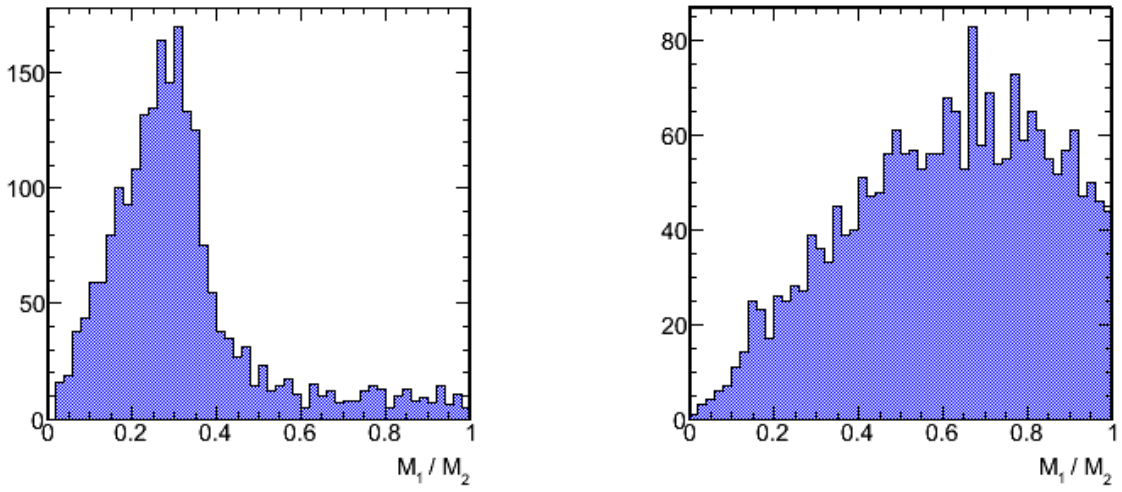
3.2 Creating and testing the BNN

Given our five input variables (two values of $\cos\theta_T$, two values of Δp_T, and the ratio M_1/M_2), our next task is to build a discriminant function with these variables. To do so, we looped over all of the events in our data file, calculated all five variables for both the correct and incorrect pairings, and wrote these values to two text files (one for the correct values, and one for the incorrect values). Again, as before, we mixed the data files into one large data file, and ran the BNN software with this file as the input.

As a test, we evaluated the resulting discriminant function $D(\vec{x})$ on a second simulated signal file of the same format. The resulting histogram of the discriminant is shown in Fig. 3.6. This result is quite similar to the general trend that we hoped to see (Fig. 2.1): the discriminant evaluates to nearly 1 most of the time for the correct pairing, and nearly 0 most of the time for the incorrect pairing.

In practice, one might use this discriminant as follows. For a given set of four leptons, pick one of the two possible pairings, calculate the five input variables, and evaluate the discriminant function. Then repeat for the other pairing. Whichever pairing has the higher value of the discriminant is taken to be the correct pairing.

Given this algorithm, we can test how often this algorithm gets the correct result. Remember, our data files order the leptons such that we know which leptons came from which Z boson; we already know the correct pairings. In order to test the discriminant, we simply evaluate the discriminant for the correct pairing, and again for the incorrect pairing. We then count how often the discriminant function has a larger value for the correct pairing, as it should. When this calculation was performed on a second data file (different from the



(a) Correct pairing

(b) Incorrect pairing

Figure 3.5: Histograms of invariant mass ratio for correct and incorrect pairings

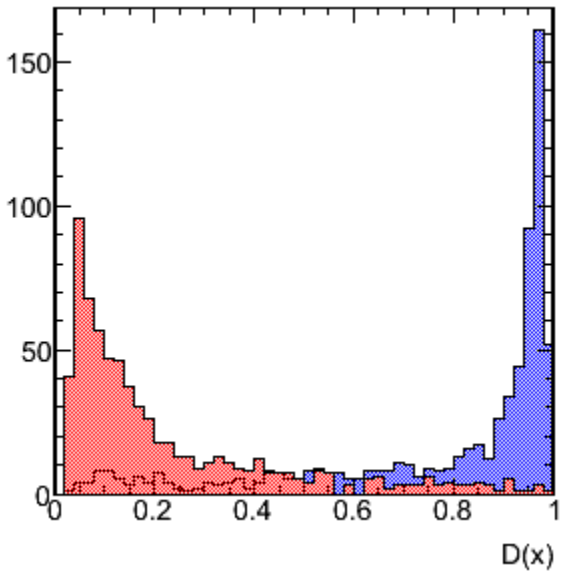


Figure 3.6: Histogram of discriminant function for correct pairing (blue) and incorrect pairing (red).

one used to train the discriminant), the resulting efficiency that we obtained is 88.6%; that is, the discriminant is larger for the correct pairing than it is for the incorrect pairing 88.6% of the time.

3.3 Comparison of our discriminant to previous lepton assignment algorithm

This discriminant-based solution to the lepton assignment problem is not the first attempted solution. The lepton assignment problem was previously approached using a series of computations involving numerous loops over particles and various criteria and cuts on the results. That algorithm is presently used for the official $H \to ZZ \to 4\ell$ analysis; for lack of a better name, we will refer to it as the "official" algorithm. Although the discriminant function is computationally intensive to build initially, the resulting function is much simpler to use than the official algorithm: simply call the function and obtain the numerical result. Despite the computational advantages to using this discriminant function, we must first ensure that its efficiency is comparable to that of the official algorithm.

3.3.1 The official algorithm

The selection criteria for the official algorithm [26] are more advanced when applied to reconstruction-level data or real data. When applied to our generator-level data, the rules are somewhat simplified.

First, loop over all four leptons. For two leptons ℓ_i and ℓ_j of the same flavor and opposite charge, find the invariant mass of the pair, m_{ij}. Whichever leptons give m_{ij} closest to the nominal Z boson mass (91.2 GeV) are taken to be the leptons corresponding to Z_1. The value of m_{Z_1} must be between 40 and 120 GeV. The remaining two leptons should have the same flavor and opposite charge. If so, and if their invariant mass is between 4 and 120 GeV, those leptons are said to correspond to Z_2. Additionally, one of the four leptons must have p_T greater than 20 GeV, and another must have p_T greater than 10 GeV. Events that cannot fulfill these criteria are discarded.

Note that this algorithm was only evaluated on events in which all four leptons were of the same flavor. Due to the nature of the selection rules applied, the official algorithm will

be correct every single time for events in which the pairs of leptons have different flavors. Indeed, as discussed previously, there is no ambiguity in events with leptons of different flavors; lepton assignment is only a problem when all four leptons are of the same flavor.

However, when we trained the discriminant function, in order to have a higher sample size, we used data from all available events — including events containing leptons of different flavors. For future work, it may be desirable to retrain the discriminant using only events with leptons that are all of the same flavor; this may increase the efficiency of the discriminant.

3.3.2 Results

Upon implementing these rules, the official algorithm provided the correct answer for the first data file 93.4% of the time. By comparison, the discriminant provided the correct answer 90.4% of the time. Note that this data file was the one used to train the discriminant. When tested on a second file, the official algorithm was correct 91.7% of the time, and the discriminant was correct 88.6% of the time. Although our discriminant-based algorithm is correct slightly less often than the official algorithm, our algorithm and the official algorithm perform with roughly equal efficiencies.

3.4 Beyond the lepton assignment problem: How these results are used

Recall that the ultimate goal of our analysis is not merely to solve the lepton assignment problem. This solution is simply a means to an end: to eliminate background events in data from the LHC. To do so requires additional analysis. As seen in Chapter 2, we can use a BNN to build a discriminant function that distinguishes between signal and background events. However, instead of using only the Z boson masses as inputs, as we did, the current analysis uses a BNN discriminant with seven input variables.

Upon solving the lepton assignment problem, the algorithm (whether it be our algorithm or the previous, official algorithm) passes on what it presumes to be the leptons corresponding to the two Z bosons. The masses of these two Z bosons, as well as five angular variables for the leptons (referred to as the *MELA variables*), form the seven input variables of the

BNN. In much the same way that we solved the lepton assignment problem, the BNN is trained with these seven variables (using reconstruction-level data) to create a discriminant function. That function is then applied to real data obtained from the LHC, eliminating background events and increasing the statistical significance of signal observations.

3.5 Suggestions for future improvements and modifications

As stated previously, our discriminant function was trained using all available generator-level data, including data with lepton pairs of different flavors. However, in those cases, lepton assignment is not a problem: due to conservation of flavor, we can immediately state the correct lepton assignment *a priori*. We may consider retraining the discriminant function using only cases in which all four leptons are of the same flavor.

Additionally, the sample size used to train this discriminant could be increased. The data file that we used to train the discriminant contains 2,165 unweighted events. We also have access to three other data files that are all of the same format. It is possible to use ROOT to link together multiple data files, extracting data from them as though they were simply one large file. We recommend using this technique to increase the size of the training sample for the BNN, which should increase the discriminant's efficiency.

However, because this discriminant will eventually be applied to real data, the next logical step would be to retrain the discriminant using reconstruction-level data. This will require some modifications. For example, this will likely involve applying selection criteria similar to those currently used by the official algorithm. Additionally, there are occasionally events in the data in which there are more than four leptons. In principle, additional leptons simply provide additional possible pairings, beyond the two possibilities considered for four leptons. If we apply the discriminant function to all possible pairings, the pairing with the highest value of the discriminant should still be the correct pairing. This scenario should be investigated further, in case further modifications are necessary.

CHAPTER 4

CONCLUSION

The analysis of this Higgs decay channel is computationally quite involved. The lepton assignment problem is but one step in this analysis. Although the lepton assignment problem has been previously solved (with an efficiency of roughly 90%, as we showed), the official algorithm currently in use is rather complicated.

We investigated a discriminant-based algorithm as a means to simplify the analysis; in principle, our BNN discriminant function can be called with a single line of code. Given this reduction in the complexity of the analysis code, as well as the fact that our algorithm performs with similar (albeit slightly reduced) efficiency, we recommend the use of this discriminant-based solution to the lepton assignment in future analyses of the $H \to ZZ \to 4\ell$ decay channel. We also suggest various ways by which the discriminant could be improved.

BIBLIOGRAPHY

1. L. Lederman and D. Teresi, *The God Particle: If the Universe Is the Answer, What Is the Question?*. (Houghton Mifflin, Boston, 1993), pp. 35-40.

2. L. Lederman and D. Teresi, *The God Particle: If the Universe Is the Answer, What Is the Question?*. (Houghton Mifflin, Boston, 1993), pp. 110-114.

3. R.A. Serway, C.J. Moses, and C.A. Moyer, *Modern Physics*, 3rd ed. (Thomson, Belmont, 2005), pp. 108-125, 466-467.

4. R.A. Serway, C.J. Moses, and C.A. Moyer, *Modern Physics*, 3rd ed. (Thomson, Belmont, 2005), pp. 547-548, 574-581.

5. D.J. Griffiths, *Introduction to Elementary Particles*, 1st ed. (Wiley, New York, 1987), pp. 17-20, 46-48.

6. Particle Data Group, "Gauge and Higgs Bosons," Phys. Lett. B **667** (1), 385-476 (2008).

7. F. Close, *Particle Physics: A Very Short Introduction* (Oxford, Oxford, 2004), p. 124.

8. K Nakamura et al. (Particle Data Group), J. Phys. G: Nucl. Part. Phys. **37**, 075021 (2010).

9. C.R. Nave, "Particle lifetimes from the uncertainty principle," *HyperPhysics*, 2013, http://hyperphysics.phy-astr.gsu.edu/hbase/quantum/parlif.html (30 Jan 2013).

10. LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions," *High Energy Physics - Phenomenology*, 15 Jan 2012, `http://arxiv.org/abs/1201.3084` (30 Jan 2013).

11. CMS Collaboration, Phys. Lett. B **716**, 30-61 (2012).

12. ALEPH, DELPHI, L3, OPAL Collaborations, and LEP Working Group for Higgs Boson Searches, Phys. Lett. B **565**, 61 (2003).

13. M.W. Grunewald, "Precision Electroweak Measurements and Constraints on the Standard Model," 13 Dec 2010, http://cds.cern.ch/record/1313716 (28 Jan 2013).

14. CERN, "The Z Factory," 2008, http://public.web.cern.ch/PUBLIC/en/Research/LEP-en.html (28 Jan 2013).

15. M.M. Kado and C.G. Tully, Annu. Rev. Nucl. Part. Sci. **52** (1), 65-113 (2002).

16. Fermilab, "Energy Frontier," 17 Oct 2012, http://www.fnal.gov/pub/science/experiments/energy/tevatron/ (28 Jan 2013).

17. Fermilab, "Fermilab Today," 10 Sept 2010, http://www.fnal.gov/pub/today/archive/archive_2010/today10-09-10.html (28 Jan 2013).

18. L. Evans and P. Bryant, JINST **3**, S08001 (2008).

19. CERN, "CERN experiments observe particle consistent with long-sought Higgs boson," *CERN press office*, 4 July 2012, http://press.web.cern.ch/press-releases/2012/07/cern-experiments-observe-particle-consistent-long-sought-higgs-boson (28 Jan 2013).

20. ATLAS Collaboration, Phys. Lett. B **716**, 1-29 (2012).

21. C. Wong, *Introduction to High-Energy Heavy-Ion Collisions* (World Scientific, Singapore, 1994), p. 24.

22. S.R. Saucedo, M.S. thesis, Florida State University, 2007.

23. CERN, "A Gentle introduction," *Discovering ROOT*, 2013, http://root.cern.ch/drupal/content/gentle-introduction (30 Jan 2013).

24. H.B. Prosper, D. Kau, S. Jain, *et al.*, "A Second Look at Bayesian Neural Networks in the Search for Single Top Quarks in 1 fb^{-1} of Data," DØNote 5361 (2007).

25. S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, Phys. Lett. B **195** (2), 216-222 (1987).

26. CMS Collaboration, "Updated results on the new boson discovered in the search for the standard model Higgs boson in the ZZ to 4 leptons channel in pp collisions at sqrt(s) = 7 and 8 TeV," 16 Nov 2012, http://cds.cern.ch/record/1494488?ln=en (12 Feb 2013).

BIOGRAPHICAL SKETCH

Daniel Charles was born on May 14, 1988. He graduated from F.W. Springstead High School in Spring Hill, FL in 2006 and began college at Florida State University that fall. After receiving his A.A. degree in 2007, he transferred to the University of South Florida (USF). He participated in the University of Florida's Research Experience for Undergraduates program in physics during the summer of 2009, working with Professor David Tanner. During the summer of 2010, he participated in the Princeton Plasma Physics Laboratory's National Undergraduate Fellowship program in plasma physics, working under Dr. Paul Woskov at the Massachusetts Institute of Technology (MIT). He completed his B.S. degree in Physics at USF in 2011, and entered FSU's graduate program in physics that fall, where he is currently working to complete his M.S. degree in physics.