Instituto Politecnico Nacional

Centro de Investigación en Computación
Laboratorio de Ciencia de Datos

Desenredamiento de parámetros latentes

T E S I S

para obtener el título de:

Maestría en Ciencias de la Computación

presenta:

José de Jesus Daniel Aguirre Arzate

Tutores:

Dr. Ricardo Menchaca Méndez

Dr. Lukas Nellen Filla

México, CDMX
11 de octubre de 2021

Declaración de Autoría

I, José de Jesus Daniel AGUIRRE ARZATE, declare that this thesis titled, «Desenredamiento de parámetros latentes» and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

«Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.»

Dave Barry

INSTITUTO POLITECNICO NACIONAL

Resumen

CIC
Centro de Investigación en Computación

Maestría en Ciencias de la Computación

Desenredamiento de parámetros latentes

por José de Jesus Daniel AGUIRRE ARZATE

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too…

Agradecimientos

The acknowledgments and the people to thank go here, don't forget to include your project advisor…

Índice general

Índice de figuras

Índice de cuadros

Índice de Abreviaturas

LAH	**L**ist **A**bbreviations **H**ere
WSF	**W**hat (it) **S**tands **F**or

Constantes Físicas

Speed of Light	$c_0 = 2.997\,924\,58 \times 10^8\,\mathrm{m\,s^{-1}}$ (exact)

Índice de Símbolos

a	distance	m
P	power	W ($\mathrm{J\,s^{-1}}$)
ω	angular frequency	rad

For/Dedicated to/To my…

Capítulo 1

Introduccion

En este capitulo se muestra la principal motivacion para el desarrollo de esta tesis, asi como una breve perspectiva de la problematica general. Como capitulo introductorio este contiene la introduccion, el planteamiento del problema, el objetivo, las fronteras del estudio y la estructura del escrito.

1.1. Introduccion

En el area de la fisica de altas energias (HEP) las tecnicas de aprendizaje maquina (ML) siempre estuvieron presentes. Debido a la sorprendente efectividad de tecnicas modernas como el aprendizaje profundo, se comenzaron a adaptar y desarrollar estos metodos en todos los rubros del campo. Algunas de las aplicaciones van desde los enfoques que se tienen en la parte experimental, la fenomenologica o en el anlisis teorico de los eventos.

En los experimentos mas importantes del campo, el tratamiento y analisis de datos es una tarea fundamental. Tecnicas como arboles de desicion, maquinas de soporte vectorial, algoritmos geneticos, entre otras, fallaban cuando la dimensionalidad de los datos aumentaba. Como referencia de la alta dimensionalidad, en el gran colisionador de hadrones LHC, las colisiones ocurren con una frecuencia de aproximadamente 40Mhz, ademas de que cada colision genera un gran numero de particulas y en particular el LHC tiene alrededor $O(10^8)$ sensores para su deteccion.

Debido a que las observaciones son fundamentalmente probabilisticas se tiene un modelo estadistico que describe la probabilidad de observar un evento dado los parametros de una teoria. Pero la alta dimensionalidad, junto con los grandes volumenes de datos generan un problema, ya que el modelo de los datos experimentales no se conoce explicitamente. Sin embargo, si se tiene acceso a muestras de datos generados por simuladores estocasticos que modelan la fisica de las interacciones. Herramientas como PYTHIA, HERWING, GEANT, CORSIKA se les suele denominar como simuladores de Monte Carlo, los cuales cumplen con dos necesidades, la primera es aproximar el modelo estadistico al muestrar de un espacio enorme de procesos no observados o latentes y la segunda es generar una base de datos.

Entre las tareas de bajo nivel se tiene la identificacion de particulas y la reconstruccion de la energia/momento de la particula en cuestion. Debido a que los simuladores completos que describen las interacciones de las particulas con la materia, son computacionalmente intensos y se llevan gran parte del presupuesto computacional de las colaboraciones, los simuladores rapidos son escenciales. Simuladores como GEANT y CORSIKA que generan una excelente descripcion de interacciones hadronicas son lentos. En los ultimos anos ha nacido un gran interes por usar redes neuronales generativas para aumentar la velocidad de las simulaciones y tal vez llegar a usar estos metodos directamente en datos generados por colisiones reales y hacer tunning en en el momento.

El presente trabajo esta fundamentado en el desarrollo de algoritmos de aprendizaje maquina profundo, especificamente reges generativas GAN para la generacion de interacciones hadronicas. Esto debido a la necesidad de generar simulaciones precisas y de una manera mas rapida, ya que actualizaciones a dichos experimentos como el de Alta Luminosidad al

LHC exigiran una mayor capacidad computacional que no se tiene con la proyeccion de presupuestos actuales. Existe un gran interes por parte de la comunidad en usar metodos de aprendizaje no supervisado como GANs o VAEs para generar espacios de caracteristicas con una dimensionalidad alta. Uno de los mayores desafios que hay al usar estos metodos es como cuantificar su desempeno.

1.2. Plantamiento del problema

El planteamineto se fundamenta en lo siguiente:
Sera posible disenar un metodo que utilize redes neuronales generativas que logre simular cascadas hadronicas precisas y a su vez la aquitectura no mezcle sus paramatros latentes que podrian estar asociados a la energia y momento de la particula.

1.3. Objetivo de la tesis

Objetivo principal:

- Disenar y implementar una red neuronal generativa que no mezcle sus parametros latentes para generar cascadas hadornicas acordes a simulaciones obtenidas por el software CORSIKA.

Objetivos particulares:

I. Disenar una arquitectura que integre algunos metodos de aprendizaje no supervisado para la simulacion de interacciones hadronicas.

II. Entrenar esa arquitectura con una base de datos generada medienta el software CORSIKA.

III. Mostrar el algoritmo que genera las simulaciones y comparar con simulaciones generadas mediante CORSIKA.

1.4. Delimitacion del tema

1.5. Contribucion de la tesis

1.6. Organizacion de la tesis

Capítulo 2

Estado del arte

2.1. Antecedentes

La fisica de particulas se encarga de estudiar los constituyentes subatomicos de la materia para asi poder responder a preguntas fundamentales como, Cuantas particulas hay?, Cuales son sus propiedades? y Como interactuan?. Responder a estas preguntas llevaran a direcciones favorables para la solucion de problemas como la naturaleza de la masa y de la antimateria, la dimensionalidad del espacio, la unificacion de las fuerzas fundamentales y el refinamiento del modelo estandar. Para comenzar a responder estas preguntas se tienen dos enfoques basicos, el teorico y el experimental. Por el lado teorico hay limitantes en la teoria actual y para probar la validez de otras teorias los experimentos son esenciales.

Uno de los experimentos mas importantes que reflejan el estado del arte del campo es el Gran Colisionador de Hadornes (LHC), el cual tuvo uno de sus mayores logros al encontrar el Boson de Higgs en el 2012.

Schwartz2021 La fisica de particulas estudia los constituyentes dsubatomicos de la materia, preguntas como: Cuantas hay? Que propiedades tienen? Como interactuan? Hay dos maneras basicas de comenzar a responder estas preguntas: una teorica y otra experimental. Pero por el lado teorico hay limitantes en cunato a la teoria establecida. para probar alguna de las otras teorias los experimentos son esenciales. Uno de los experimentos mas importantes es el LHC el cual uno de sus mayores logros fue encontrar el Boson de Higgs en 2012. usualmente las particulas generadas en las colisiones del LHC viven por fraciciones de segundo. (la vida del Boson de Higgs es de $10^{-22}s$ esto hace que el arte de la fisica de particulas experimental moderna involucre encontrar diciadores de que una particula fue creada a pesar de que nunca la vimos. Por ejemplo solo una en un billos de colisiones de protones en el LHC produce un Boson de Higgs y solo uno de cada 10000 de estos son faciles de ver. Encontrar nuevas particulas es como intantar encontrar un tipo particular de paja en un pajar. Afortunadamente los problemas del tipo paja en pajar son en los que el aprendizaje maquina es excelente resolviendo.

Hay dos aspectos de los problemas de HEP que los ahcen unicos comparados con otros campos donde el AM se aplica. La primera es que la fisica de paprticulas esta fovernada por la mecanica cuantica. AAsi como el gato de schridinger una colision en el LHC puede producir un boson de higgs o no producirlo al mismo tiempo.

Bourilkov2019 La manera tradicional de analizar datos o generar datos es desarrollar algoritmos basados en conocimiento, despues implementarlos en software y usar los programas resultants para analisis o generacion. Ete proceso es complejo y analizar datasets complejos con muchas inputs variables se vuelve una tarea intractable. El AM atacan estre problema de una manera diferente, en vez de que los humanos creen estos algoritmos altamente especializados, los algoritmos aprenden de los datos para construir el modelo. Estos modelos pueden predecir el comportamiento de nuevos datos, para detectar anomalias o generar datos simulados.

Con los metodos tradicionales de analisis, la fisica ha avanzado rapidamente, estableciendo el Modelo Estandar de la fisica de particulas y su analogo cosmologico Lambda CDM. En los anios venideros aumentara el volumen de datos y la complejidad de analisis en experimentos como el LHC por lo tanto extraer la fisica de fondo usando los metodos tradicionales se vuelve mas complejo o siemplemente imposbile en tiempos razonables.

Las primeras aplicaciones de AM en HEP usualmente usabana arboles de decisiones: un modelo tipo arbol para tomar decisiones, comenzando en la raiz, subiendo hasta las hojas donde cada hoja representa una desicion. En el campo las tecnicas mas usadas son BDT Booosted Desicion Trees que convierten weak to strong learners.

Albertsson2019 Principales objetivos en la era post boson de Higgs: Aprovecahr al maximo todo el potencial del LHC y su actualizacion de alta luminosidad HL LHC para exoerimentos de neutrinos actuales y futuros. La actualizacion HL LHC integrara datos que son 20 mayores que los datos actulaes que produce el LHC. Trayendo consigo nuevos desafios cualitativos y cunatitativos debido a los tamanos de los eventos, volumenes de datos y complejidad. El alcance de os experimentos estara limitado por el desempeno de los algoritmos y los recursos computacionales.

Para incorporar tecnicas de aprnedizaje maquina en flujos de trabajo de HEP se requeriran avances en la investigacion y el desarrollo en los proximos 5 anios. Algunas areas donde se necesitan mejoras significativas son: Desempenio en los algoritmos de reconstruccion y analsis. Tiempo de ejecucion de partes computacionalmente expensive de simulacion de eventos, reconocimiento de patrones y calibracion Implementacion en tiempo real de algortimos de AM Reduccion de la huella de los datos, usando comprension de datos, ubicacion y acceso.

Los principales objetivos de los experimentos dentro de HEP van de la mano uno ocn el otro, probar el modelo estandar con mayor presicion y buscar nuevas particulas que el modelo estandar no contempla. Ambas tareas implican la identificacion de senales extranas en fondos de ruido immenso. Al aumentar el numero de colisiones observadas con la actualizacion HL LHC hace que lo anterior se vuelva un gran desafio.

Los algoritmos de AM que actualmente se usan con mayor frecuencia en el campo HEP son Boosted Decision Trees (BDTs) y redes neuronaless (NNs). Tipicamente se seleccionan als variables relevantes a la fisica y el modelo de AM se entrena para clasificar o hacer regresion usando senales y eventos de fondo (instancias). Entrenar el modelo es el paso que conlleva mas trabajo tanto humano como de los recursos computacionales, mientras que el paso de inferencia es relativamente barato. BDTs y NNs son usados para clasificar particulas y eventos. Para regresion se usa para obtener el mejor estimado de la energia de una particula basado en los measurements de varios detectores. Delas arquitecturas de NNs mas usadas en HEP se tienen las FCN, CNN y RNN. Adicionando modelos generativos como GANs (recientemente) o VAEs. La gran mayoria de algoritmos de AM se usan para analisis de series de tiempo. En general no son relevantes para el analisis de datos de HEP ya que lso eventos son independiewntes uno del otro.

Guest2018 En la fisica de particulas el AM tiene muchas aplicaciones ya que muchas tareas implican clasificacion en espacios de alta dimensionalidad. En los niveles mas bajos la herramientas de AM pueden reconstruir o encontrar tracks en detectores individuales. Usando la informacion de varios detectores y identificacion de objetos podemos identificar eelctrones, protones o tau leptones. Estas herramientas tambien son ampliamente usadas para clasificacion de eventos como background like o signal like tanto al final del analisis estadistico y en el trigger incial. En aplicaciones de alto nivel hay trabajos en la busqueda del t quark, la busqueda del Boson de Higgs y el descubrimiento del Boson de Higgs.

2.2. Aprendizaje unsupervisado

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

2.2.1. Gans

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

2.2.2. Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

2.3. Main Section 2

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Capítulo 3

Redes Generativas

3.1. Introduccion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

3.1.1. Fundamento teorico

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

3.2. Lineas de investigacion

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

Capítulo 4

Metodo experimental

4.1. Preprocesamiento de datos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

4.2. Algoritmos

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

4.2.1. Subsection 1

Nunc posuere quam at lectus tristique eu ultrices augue venenatis. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam erat volutpat. Vivamus sodales tortor eget quam adipiscing in vulputate ante ullamcorper. Sed eros ante, lacinia et sollicitudin et, aliquam sit amet augue. In hac habitasse platea dictumst.

4.2.2. Subsection 2

Morbi rutrum odio eget arcu adipiscing sodales. Aenean et purus a est pulvinar pellentesque. Cras in elit neque, quis varius elit. Phasellus fringilla, nibh eu tempus venenatis, dolor elit posuere quam, quis adipiscing urna leo nec orci. Sed nec nulla auctor odio aliquet consequat. Ut nec nulla in ante ullamcorper aliquam at sed dolor. Phasellus fermentum magna in augue gravida cursus. Cras sed pretium lorem. Pellentesque eget ornare odio. Proin accumsan, massa viverra cursus pharetra, ipsum nisi lobortis velit, a malesuada dolor lorem eu neque.

4.3. Desarrollo del metodo

4.4. Mapa del metodo experimental

Capítulo 5

Resultados

5.1. Dsicusion de resultados obtenidos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

Capítulo 6

Conlusiones

6.1. Contribuciones y conclusiones, puntuales, obtenidas

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam ultricies lacinia euismod. Nam tempus risus in dolor rhoncus in interdum enim tincidunt. Donec vel nunc neque. In condimentum ullamcorper quam non consequat. Fusce sagittis tempor feugiat. Fusce magna erat, molestie eu convallis ut, tempus sed arcu. Quisque molestie, ante a tincidunt ullamcorper, sapien enim dignissim lacus, in semper nibh erat lobortis purus. Integer dapibus ligula ac risus convallis pellentesque.

6.2. Publicaciones

Sed ullamcorper quam eu nisl interdum at interdum enim egestas. Aliquam placerat justo sed lectus lobortis ut porta nisl porttitor. Vestibulum mi dolor, lacinia molestie gravida at, tempus vitae ligula. Donec eget quam sapien, in viverra eros. Donec pellentesque justo a massa fringilla non vestibulum metus vestibulum. Vestibulum in orci quis felis tempor lacinia. Vivamus ornare ultrices facilisis. Ut hendrerit volutpat vulputate. Morbi condimentum venenatis augue, id porta ipsum vulputate in. Curabitur luctus tempus justo. Vestibulum risus lectus, adipiscing nec condimentum quis, condimentum nec nisl. Aliquam dictum sagittis velit sed iaculis. Morbi tristique augue sit amet nulla pulvinar id facilisis ligula mollis. Nam elit libero, tincidunt ut aliquam at, molestie in quam. Aenean rhoncus vehicula hendrerit.

6.3. Aplicacion y extension generadas del trabajo

6.4. Trabajos futuros

Capítulo 7

LaTex Review

7.1. Welcome and Thank You

Welcome to this LaTeX Thesis Template, a beautiful and easy to use template for writing a thesis using the LaTeX typesetting system.

If you are writing a thesis (or will be in the future) and its subject is technical or mathematical (though it doesn't have to be), then creating it in LaTeX is highly recommended as a way to make sure you can just get down to the essential writing without having to worry over formatting or wasting time arguing with your word processor.

LaTeX is easily able to professionally typeset documents that run to hundreds or thousands of pages long. With simple mark-up commands, it automatically sets out the table of contents, margins, page headers and footers and keeps the formatting consistent and beautiful. One of its main strengths is the way it can easily typeset mathematics, even *heavy* mathematics. Even if those equations are the most horribly twisted and most difficult mathematical problems that can only be solved on a super-computer, you can at least count on LaTeX to make them look stunning.

7.2. Learning LaTeX

LaTeX is not a WYSIWYG (What You See is What You Get) program, unlike word processors such as Microsoft Word or Apple's Pages. Instead, a document written for LaTeX is actually a simple, plain text file that contains *no formatting*. You tell LaTeX how you want the formatting in the finished document by writing in simple commands amongst the text, for example, if I want to use *italic text for emphasis*, I write the `\emph{text}` command and put the text I want in italics in between the curly braces. This means that LaTeX is a «mark-up» language, very much like HTML.

```
from sys import argv

def reader(file):
  """
  This funciton returns a matrix M[i][j] of the input file.csv.
  """
  lines = open(file, 'r').readlines()
  a = []
  for i in range(0, len(lines)):
    a.append(lines[i].strip().split(','))
  return a

def reader_grammar(file):
  lines = open(file, 'r').readlines()
  a = []
  for i in range(0, len(lines)):
    a.append(lines[i].strip().split('->'))
  return a

def states(matr_of_file):
  """
  This function returns a list of the sates of a given transition matrix of a
    machine.
  The input must be a list of list of the form M[i][j].
  """
```

```python
  a = []
  for i in range(1,len(matr_of_file)):
    a.append(matr_of_file[i][0])
  return a

def alphabet(matr_of_file):
  """
  This function returns a list of the alphabet of the machine M given by the input in matrix form M[i][j].
  """
  a = []
  for i in range(1, len(matr_of_file[0])-1):
    a.append(matr_of_file[0][i])
  return a

def number_list(l):
  """
  This function returns a list of number from 1 to len(l) given the list l.
  """
  a = []
  for x in range(1,len(l)+1):
    a.append(x)
  return a

def trans_func(matr_of_file, state, char):
  """
  This function returns the state that the machine goes to, given an state and a character of the alphabet.
  e.g. trans_funciton('state[0]', 'a') = 'state1'
  """
  return matr_of_file[dict(zip(states(matr_of_file), number_list(states(matr_of_file))))[state]][dict(zip(alphabet(matr_of_file), number_list(alphabet(matr_of_file))))[char]]

def init_state(matr_of_file):
  """
  This function returns te initial state of the machine M given by the input in matrix form M[i][j].
  """
  return matr_of_file[1][0]

def eof(matr_of_file):
  """
  This function returns a list of the acceptance states of the machine M given by the input in matrix form M[i][j].
  """
  a = []
  for i in range(0,len(states(matr_of_file))):
    if matr_of_file[i+1][len(matr_of_file[0])-1] == 'accept':
      a.append(states(matr_of_file)[i])
  return a

def isinalpha(a, alph):
  """
  This is a boolean function that tells if the character a is in the alphabet alph.
  e.g. isinalpha('a', ['a', 'b']) = True
  """
  return a in alph

def machine(file, string):
  """
  This function implements the SFA give the input file in matrix form M[i][j] and given a string.
  e.g. machine(reader('filename.csv'), 'hola') = No
  """
  if isinalpha(string[0], alphabet(file)):
    s0 = trans_func(file, init_state(file), string[0])
    i = 0
    while i < len(string)-1:
      n = s0
      i += 1
      if isinalpha(string[i], alphabet(file)):
```

```
        try:
          s0 = trans_func(file, n, string[i])
        except:
          return 'no'
      else:
        return 'no'
    if (s0 in eof(file)) == True:
      return 'yes'
    else:
      return 'no'
  else:
    return 'no'

scrpt, file, string = argv
print(machine(reader(file), string))
```

7.2.1. A (not so short) Introduction to LaTeX

If you are new to LaTeX, there is a very good eBook – freely available online as a PDF file – called, «The Not So Short Introduction to LaTeX». The book's title is typically shortened to just *lshort*. You can download the latest version (as it is occasionally updated) from here: `http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf`

It is also available in several other languages. Find yours from the list on this page: `http://www.ctan.org/tex-archive/info/lshort/`

It is recommended to take a little time out to learn how to use LaTeX by creating several, small 'test' documents, or having a close look at several templates on:
`http://www.LaTeXTemplates.com`
Making the effort now means you're not stuck learning the system when what you *really* need to be doing is writing your thesis.

7.2.2. A Short Math Guide for LaTeX

If you are writing a technical or mathematical thesis, then you may want to read the document by the AMS (American Mathematical Society) called, «A Short Math Guide for LaTeX». It can be found online here: `http://www.ams.org/tex/amslatex.html` under the «Additional Documentation» section towards the bottom of the page.

7.2.3. Common LaTeX Math Symbols

There are a multitude of mathematical symbols available for LaTeX and it would take a great effort to learn the commands for them all. The most common ones you are likely to use are shown on this page: `http://www.sunilpatel.co.uk/latex-type/latex-math-symbols/`

You can use this page as a reference or crib sheet, the symbols are rendered as large, high quality images so you can quickly find the LaTeX command for the symbol you need.

7.2.4. LaTeX on a Mac

The LaTeX distribution is available for many systems including Windows, Linux and Mac OS X. The package for OS X is called MacTeX and it contains all the applications you need – bundled together and pre-customized – for a fully working LaTeX environment and work flow.

MacTeX includes a custom dedicated LaTeX editor called TeXShop for writing your '`.tex`' files and BibDesk: a program to manage your references and create your bibliography section just as easily as managing songs and creating playlists in iTunes.

7.3. Getting Started with this Template

If you are familiar with LaTeX, then you should explore the directory structure of the template and then proceed to place your own information into the *THESIS INFORMATION* block of the `main.tex` file. You can then modify the rest of this file to your unique specifications based on your degree/university. Section 7.5 on page 19 will help you do this. Make sure you also read section 7.7 about thesis conventions to get the most out of this template.

If you are new to LaTeX it is recommended that you carry on reading through the rest of the information in this document.

Before you begin using this template you should ensure that its style complies with the thesis style guidelines imposed by your institution. In most cases this template style and layout will be suitable. If it is not, it may only require a small change to bring the template in line with your institution's recommendations. These modifications will need to be done on the `MastersDoctoralThesis.cls` file.

7.3.1. About this Template

This LaTeX Thesis Template is originally based and created around a LaTeX style file created by Steve R. Gunn from the University of Southampton (UK), department of Electronics and Computer Science. You can find his original thesis style file at his site, here: `http://www.ecs.soton.ac.uk/~srg/softwaretools/document/templates/`

Steve's `ecsthesis.cls` was then taken by Sunil Patel who modified it by creating a skeleton framework and folder structure to place the thesis files in. The resulting template can be found on Sunil's site here: `http://www.sunilpatel.co.uk/thesis-template`

Sunil's template was made available through `http://www.LaTeXTemplates.com` where it was modified many times based on user requests and questions. Version 2.0 and onwards of this template represents a major modification to Sunil's template and is, in fact, hardly recognisable. The work to make version 2.0 possible was carried out by Vel and Johannes Böttcher.

7.4. What this Template Includes

7.4.1. Folders

This template comes as a single zip file that expands out to several files and folders. The folder names are mostly self-explanatory:

Appendices – this is the folder where you put the appendices. Each appendix should go into its own separate `.tex` file. An example and template are included in the directory.

Chapters – this is the folder where you put the thesis chapters. A thesis usually has about six chapters, though there is no hard rule on this. Each chapter should go in its own separate `.tex` file and they can be split as:

- Chapter 1: Introduction to the thesis topic
- Chapter 2: Background information and theory
- Chapter 3: (Laboratory) experimental setup
- Chapter 4: Details of experiment 1
- Chapter 5: Details of experiment 2
- Chapter 6: Discussion of the experimental results
- Chapter 7: Conclusion and future directions

This chapter layout is specialised for the experimental sciences, your discipline may be different.

Figures – this folder contains all figures for the thesis. These are the final images that will go into the thesis document.

7.4.2. Files

Included are also several files, most of them are plain text and you can see their contents in a text editor. After initial compilation, you will see that more auxiliary files are created by LaTeX or BibTeX and which you don't need to delete or worry about:

example.bib – this is an important file that contains all the bibliographic information and references that you will be citing in the thesis for use with BibTeX. You can write it manually, but there are reference manager programs available that will create and manage it for you. Bibliographies in LaTeX are a large subject and you may need to read about BibTeX before starting with this. Many modern reference managers will allow you to export your references in BibTeX format which greatly eases the amount of work you have to do.

MastersDoctoralThesis.cls – this is an important file. It is the class file that tells LaTeX how to format the thesis.

main.pdf – this is your beautifully typeset thesis (in the PDF file format) created by LaTeX. It is supplied in the PDF with the template and after you compile the template you should get an identical version.

main.tex – this is an important file. This is the file that you tell LaTeX to compile to produce your thesis as a PDF file. It contains the framework and constructs that tell LaTeX how to layout the thesis. It is heavily commented so you can read exactly what each line of code does and why it is there. After you put your own information into the *THESIS INFORMATION* block – you have now started your thesis!

Files that are *not* included, but are created by LaTeX as auxiliary files include:

main.aux – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file.

main.bbl – this is an auxiliary file generated by BibTeX, if it is deleted, BibTeX simply regenerates it when you run the `main.aux` file. Whereas the `.bib` file contains all the references you have, this `.bbl` file contains the references you have actually cited in the thesis and is used to build the bibliography section of the thesis.

main.blg – this is an auxiliary file generated by BibTeX, if it is deleted BibTeX simply regenerates it when you run the main `.aux` file.

main.lof – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It tells LaTeX how to build the *List of Figures* section.

main.log – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It contains messages from LaTeX, if you receive errors and warnings from LaTeX, they will be in this `.log` file.

main.lot – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file. It tells LaTeX how to build the *List of Tables* section.

main.out – this is an auxiliary file generated by LaTeX, if it is deleted LaTeX simply regenerates it when you run the main `.tex` file.

So from this long list, only the files with the `.bib`, `.cls` and `.tex` extensions are the most important ones. The other auxiliary files can be ignored or deleted as LaTeX and BibTeX will regenerate them.

7.5. Filling in Your Information in the `main.tex` File

You will need to personalise the thesis template and make it your own by filling in your own information. This is done by editing the `main.tex` file in a text editor or your favourite LaTeX environment.

Open the file and scroll down to the third large block titled *THESIS INFORMATION* where you can see the entries for *University Name*, *Department Name*, etc …

Fill out the information about yourself, your group and institution. You can also insert web links, if you do, make sure you use the full URL, including the `http://` for this. If you don't want these to be linked, simply remove the `\href{url}{name}` and only leave the name.

When you have done this, save the file and recompile `main.tex`. All the information you filled in should now be in the PDF, complete with web links. You can now begin your thesis proper!

7.6. The `main.tex` File Explained

The `main.tex` file contains the structure of the thesis. There are plenty of written comments that explain what pages, sections and formatting the LaTeX code is creating. Each major document element is divided into commented blocks with titles in all capitals to make it obvious what the following bit of code is doing. Initially there seems to be a lot of LaTeX code, but this is all formatting, and it has all been taken care of so you don't have to do it.

Begin by checking that your information on the title page is correct. For the thesis declaration, your institution may insist on something different than the text given. If this is the case, just replace what you see with what is required in the *DECLARATION PAGE* block.

Then comes a page which contains a funny quote. You can put your own, or quote your favourite scientist, author, person, and so on. Make sure to put the name of the person who you took the quote from.

Following this is the abstract page which summarises your work in a condensed way and can almost be used as a standalone document to describe what you have done. The text you write will cause the heading to move up so don't worry about running out of space.

Next come the acknowledgements. On this page, write about all the people who you wish to thank (not forgetting parents, partners and your advisor/supervisor).

The contents pages, list of figures and tables are all taken care of for you and do not need to be manually created or edited. The next set of pages are more likely to be optional and can be deleted since they are for a more technical thesis: insert a list of abbreviations you have used in the thesis, then a list of the physical constants and numbers you refer to and finally, a list of mathematical symbols used in any formulae. Making the effort to fill these tables means the reader has a one-stop place to refer to instead of searching the internet and references to try and find out what you meant by certain abbreviations or symbols.

The list of symbols is split into the Roman and Greek alphabets. Whereas the abbreviations and symbols ought to be listed in alphabetical order (and this is *not* done automatically for you) the list of physical constants should be grouped into similar themes.

The next page contains a one line dedication. Who will you dedicate your thesis to?

Finally, there is the block where the chapters are included. Uncomment the lines (delete the `%` character) as you write the chapters. Each chapter should be written in its own file and put into the *Chapters* folder and named `LatexChapter`, `Chapter2`, etc... Similarly for the appendices, uncomment the lines as you need them. Each appendix should go into its own file and placed in the *Appendices* folder.

After the preamble, chapters and appendices finally comes the bibliography. The bibliography style (called *`authoryear`*) is used for the bibliography and is a fully featured style that will even include links to where the referenced paper can be found online. Do not underestimate how grateful your reader will be to find that a reference to a paper is just a click away. Of course, this relies on you putting the URL information into the BibTeX file in the first place.

7.7. Thesis Features and Conventions

To get the best out of this template, there are a few conventions that you may want to follow.

One of the most important (and most difficult) things to keep track of in such a long document as a thesis is consistency. Using certain conventions and ways of doing things (such as using a Todo list) makes the job easier. Of course, all of these are optional and you can adopt your own method.

7.7.1. Printing Format

This thesis template is designed for double sided printing (i.e. content on the front and back of pages) as most theses are printed and bound this way. Switching to one sided printing is as simple as uncommenting the *`oneside`* option of the `documentclass` command at the top of the `main.tex` file. You may then wish to adjust the margins to suit specifications from your institution.

The headers for the pages contain the page number on the outer side (so it is easy to flick through to the page you want) and the chapter name on the inner side.

The text is set to 11 point by default with single line spacing, again, you can tune the text size and spacing should you want or need to using the options at the very start of `main.tex`. The spacing can be changed similarly by replacing the *singlespacing* with *onehalfspacing* or *doublespacing*.

7.7.2. Using US Letter Paper

The paper size used in the template is A4, which is the standard size in Europe. If you are using this thesis template elsewhere and particularly in the United States, then you may have to change the A4 paper size to the US Letter size. This can be done in the margins settings section in `main.tex`.

Due to the differences in the paper size, the resulting margins may be different to what you like or require (as it is common for institutions to dictate certain margin sizes). If this is the case, then the margin sizes can be tweaked by modifying the values in the same block as where you set the paper size. Now your document should be set up for US Letter paper size with suitable margins.

7.7.3. References

The `biblatex` package is used to format the bibliography and inserts references such as this one (Hawthorn, Weber y Scholten, 2001). The options used in the `main.tex` file mean that the in-text citations of references are formatted with the author(s) listed with the date of the publication. Multiple references are separated by semicolons (e.g. (Wieman y Hollberg, 1991; Hawthorn, Weber y Scholten, 2001)) and references with more than three authors only show the first author with *et al.* indicating there are more authors (e.g. (Arnold y col., 1998)). This is done automatically for you. To see how you use references, have a look at the `LatexChapter.tex` source file. Many reference managers allow you to simply drag the reference into the document as you type.

Scientific references should come *before* the punctuation mark if there is one (such as a comma or period). The same goes for footnotes[1]. You can change this but the most important thing is to keep the convention consistent throughout the thesis. Footnotes themselves should be full, descriptive sentences (beginning with a capital letter and ending with a full stop). The APA6 states: «Footnote numbers should be superscripted, [...], following any punctuation mark except a dash.» The Chicago manual of style states: «A note number should be placed at the end of a sentence or clause. The number follows any punctuation mark except the dash, which it precedes. It follows a closing parenthesis.»

The bibliography is typeset with references listed in alphabetical order by the first author's last name. This is similar to the APA referencing style. To see how LaTeX typesets the bibliography, have a look at the very end of this document (or just click on the reference number links in in-text citations).

A Note on bibtex

The bibtex backend used in the template by default does not correctly handle unicode character encoding (i.e. ïnternationalçharacters). You may see a warning about this in the compilation log and, if your references contain unicode characters, they may not show up correctly or at all. The solution to this is to use the biber backend instead of the outdated bibtex backend. This is done by finding this in `main.tex`: *backend=bibtex* and changing it to *backend=biber*. You will then need to delete all auxiliary BibTeX files and navigate to the template directory in your terminal (command prompt). Once there, simply type `biber main` and biber will compile your bibliography. You can then compile `main.tex` as normal and your bibliography will be updated. An alternative is to set up your LaTeX editor to compile with biber instead of bibtex, see here for how to do this for various editors.

[1]Such as this footnote, here down at the bottom of the page.

CUADRO 7.1: The effects of treatments X and Y on the four groups studied.

Groups	Treatment X	Treatment Y
1	0.2	0.8
2	0.17	0.7
3	0.24	0.75
4	0.68	0.3

7.7.4. Tables

Tables are an important way of displaying your results, below is an example table which was generated with this code:

```
\begin{table}
\caption{The effects of treatments X and Y on the four groups studied.}
\label{tab:treatments}
\centering
\begin{tabular}{l l l}
\toprule
\tabhead{Groups} & \tabhead{Treatment X} & \tabhead{Treatment Y} \\
\midrule
1 & 0.2 & 0.8\\
2 & 0.17 & 0.7\\
3 & 0.24 & 0.75\\
4 & 0.68 & 0.3\\
\bottomrule\\
\end{tabular}
\end{table}
```

You can reference tables with \ref{<label>} where the label is defined within the table environment. See LatexChapter.tex for an example of the label and citation (e.g. Table 7.1).

7.7.5. Figures

There will hopefully be many figures in your thesis (that should be placed in the *Figures* folder). The way to insert figures into your thesis is to use a code template like this:

```
\begin{figure}
\centering
\includegraphics{Figures/Electron}
\decoRule
\caption[An Electron]{An electron (artist's impression).}
\label{fig:Electron}
\end{figure}
```

Also look in the source file. Putting this code into the source file produces the picture of the electron that you can see in the figure below.

Sometimes figures don't always appear where you write them in the source. The placement depends on how much space there is on the page for the figure. Sometimes there is not enough room to fit a figure directly where it should go (in relation to the text) and so LaTeX puts it at the top of the next page. Positioning figures is the job of LaTeX and so you should only worry about making them look good!

Figures usually should have captions just in case you need to refer to them (such as in Figure 7.1). The \caption command contains two parts, the first part, inside the square brackets is the title that will appear in the *List of Figures*, and so should be short. The second part in the curly brackets should contain the longer and more descriptive caption text.

The \decoRule command is optional and simply puts an aesthetic horizontal line below the image. If you do this for one image, do it for all of them.

LaTeX is capable of using images in pdf, jpg and png format.

FIGURA 7.1: An electron (artist's impression).

7.7.6. Typesetting mathematics

If your thesis is going to contain heavy mathematical content, be sure that LaTeX will make it look beautiful, even though it won't be able to solve the equations for you.

The «Not So Short Introduction to LaTeX» (available on CTAN) should tell you everything you need to know for most cases of typesetting mathematics. If you need more information, a much more thorough mathematical guide is available from the AMS called, «A Short Math Guide to LaTeX» and can be downloaded from: `ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf`

There are many different LaTeX symbols to remember, luckily you can find the most common symbols in The Comprehensive LaTeX~Symbol List.

You can write an equation, which is automatically given an equation number by LaTeX like this:

```
\begin{equation}
E = mc^{2}
\label{eqn:Einstein}
\end{equation}
```

This will produce Einstein's famous energy-matter equivalence equation:

$$E = mc^2 \tag{7.1}$$

All equations you write (which are not in the middle of paragraph text) are automatically given equation numbers by LaTeX. If you don't want a particular equation numbered, use the unnumbered form:

```
\[ a^{2}=4 \]
```

7.8. Sectioning and Subsectioning

You should break your thesis up into nice, bite-sized sections and subsections. LaTeX automatically builds a table of Contents by looking at all the `\chapter{}`, `\section{}` and `\subsection{}` commands you write in the source.

The Table of Contents should only list the sections to three (3) levels. A `chapter{}` is level zero (0). A `\section{}` is level one (1) and so a `\subsection{}` is level two (2). In your thesis it is likely that you will even use a `subsubsection{}`, which is level three (3). The depth to which the Table of Contents is formatted is set within `MastersDoctoralThesis.cls`. If you need this changed, you can do it in `main.tex`.

7.9. In Closing

You have reached the end of this mini-guide. You can now rename or overwrite this pdf file and begin writing your own `LatexChapter.tex` and the rest of your thesis. The easy work of setting up the structure and framework has been taken care of for you. It's now your job to fill it out!

Good luck and have lots of fun!

Guide written by —
Sunil Patel: www.sunilpatel.co.uk
Vel: LaTeXTemplates.com

Apéndice A

Frequently Asked Questions

A.1. How do I change the colors of links?

The color of links can be changed to your liking using:
`\hypersetup{urlcolor=red}`, or
`\hypersetup{citecolor=green}`, or
`\hypersetup{allcolor=blue}`.
If you want to completely hide the links, you can use:
`\hypersetup{allcolors=.}`, or even better:
`\hypersetup{hidelinks}`.
If you want to have obvious links in the PDF but not the printed text, use:
`\hypersetup{colorlinks=false}`.

Bibliografía

Arnold, A. S. y col. (mar. de 1998). «A Simple Extended-Cavity Diode Laser». En: *Review of Scientific Instruments* 69.3, págs. 1236-1239. URL: `http://link.aip.org/link/?RSI/69/1236/1`.

Hawthorn, C. J., K. P. Weber y R. E. Scholten (dic. de 2001). «Littrow Configuration Tunable External Cavity Diode Laser with Fixed Direction Output Beam». En: *Review of Scientific Instruments* 72.12, págs. 4477-4479. URL: `http://link.aip.org/link/?RSI/72/4477/1`.

Wieman, Carl E. y Leo Hollberg (ene. de 1991). «Using Diode Lasers for Atomic Physics». En: *Review of Scientific Instruments* 62.1, págs. 1-20. URL: `http://link.aip.org/link/?RSI/62/1/1`.