Exploring New Physics with Neutrino Oscillation Experiments

Masters Thesis

January 27, 2023

Faculty of Physics, University of Barcelona

Gavin King

gavin@hibernate.org

Supervised by MC González García & Michele Maltoni

Abstract

Neutrino oscillation experiments set out to measure the differences in mass between the three neutrino flavors of the Standard Model of particle physics, and probe the elements of the lepton mixing (PMNS) matrix encoding the relationship between neutrino mass and interaction eigenstates. In this work, we use the results of these experiments for a different purpose: to place constraints on parameters which, if found to be nonzero, would indicate the existence of new physics (NP) beyond the Standard Model.

Neutrino flavor oscillations arise from differences in mass between the three neutrino mass eigenstates, which enter the Hamiltonian in a term inversely proportional to the energy of the propagating neutrino. In the high-energy limit, all three flavors are effectively massless, the mass splittings vanish, and the oscillation wavelength grows too large for oscillations to be detectable.

The NP effects we consider here result from introducing new terms into the Hamiltonian with a different energy dependence, either:

- independent of energy, for the case of "vector-like" interactions, or
- directly proportional to energy, for the case of "tensor-like" interactions.

Thus, the new physics would manifest as a deviation from the expected suppression of neutrino flavor oscillations at high energy. If such new physics exists, and is detectable in current experiments, we would expect to observe a contribution to the oscillation wavelength which remains constant or grows with energy.

The effects which may be represented within this generic framework include nonstandard interactions between neutrinos and matter, couplings with spacetime torsion fields, violations of Lorentz invariance or of the equivalence principle, and violations of CPT symmetry.

The 2004 work of Gonzalez Garcia & Maltoni [1] inferred limits on parameters encoding NP effects on two-flavor neutrino oscillations, using atmospheric neutrino data from Super-Kamiokande, along with data from the long baseline KEK to Kamioka (K2K) experiment.

Here, we continue to use data from atmospheric and accelerator neutrino experiments to compute upper bounds on NP parameters. We begin by reproducing the results of [1] using updated data on atmospheric neutrino oscillations, incorporating the latest data from Super-Kamiokande, together with new data from DeepCore. We then consider the NP parameters of a model with three neutrino flavors, and calculate constraints using data from particle accelerators: Tokai to Kamioka (T2K), MINOS, and NOνA.

Acknowledgements

I'm grateful to Concha for teaching me so many interesting things about neutrinos and for correcting my many misconceptions, to Michele for making his numerical code available to me, and to both for their energy and passion and for their very detailed feedback on this document.

Indeed, I'm grateful to the Faculty of Physics at UB for allowing me this belated chance to learn all these things I've wanted to understand for quite literally my whole life.

I must also thank Sonia, Vanessa, Karla, and Ariadne for their support and for putting up with me while I had my head stuck in physics for three years.

Contents

Chapter 1 || Introduction

We begin by reviewing the nature of neutrino flavors in the Standard Model, and the mathematical formalism of three-neutrino vacuum oscillations, before turning our attention to the effects of new physics on neutrino oscillations.

1.1 Neutrino flavors in the Standard Model

Neutrinos carry neither electric nor color charge, and are therefore only produced in weak interactions. Since the weak interaction couples left-handed fields, the simplest self-consistent version of the Standard Model features only left-handed neutrino fields, incorporated as the upper $I_3 = +\frac{1}{2}$ elements of the left-handed lepton doublets $L_{\ell L}$:

$$L_{\ell L} \equiv \begin{bmatrix} \nu_{\ell L} \\ \ell_L \end{bmatrix}, \qquad \ell = e^-, \mu^-, \tau^- \tag{1.1}$$

The doublet $L_{\ell L}$ has weak hypercharge $Y = 1$, and weak isospin $I = \frac{1}{2}$ so that the charged leptons are assigned electric charge $Q = I_3 + Y/2 = -1$ and the neutrinos are uncharged with $Q = 0$.

Here, the charged lepton fields ℓ_L of flavor e, μ, and τ were taken to be fields of definite mass. The neutrino fields $\nu_{\ell L}$ are labelled by the the charged lepton flavor to which they correspond: ν_e, ν_μ, ν_τ. These fields are called *weak interaction eigenstates* since they each couple to a discrete flavor of charged lepton in weak interactions. These fields are not required by this definition to be of definite mass, and, empirically, they are not the same as the *mass eigenstates*.

To complete the lepton sector we have three right-handed $I = 0$ singlets with $Q = Y/2 = -1$:

$$\ell_R, \qquad \ell = e^-, \mu^-, \tau^- \tag{1.2}$$

The right-handed leptons are of importance to us due to their role in generating mass.

1.1.1 The nature of massive neutrinos

In the Standard Model, the mass of a charged fermion f arises from a Yukawa term coupling f_L with f_R. For the charged leptons, this term is of form:

$$-\mathcal{L}_m = m_\ell \, \bar{\ell}_R \, \ell_L + \text{h.c.} \tag{1.3}$$

But it's impossible to form a Lorentz-invariant and gauge-invariant Dirac mass term using only a ν_L field, and so in the Standard Model described above, with no right-handed neutrinos, neutrinos are necessarily massless.

The experimental evidence for neutrino flavor oscillations demonstrates conclusively that this Standard Model is wrong. Neutrinos have mass, and there are at least three species of neutrino with *different* masses.

Therefore, a Standard Model consistent with experiment must have a Lagrangian with a mass term for the neutrino fields. The conventional way to achieve this is to introduce right-handed "sterile" neutrinos ν_R with no Standard Model interactions, along with:

- a Dirac mass term coupling ν_R with ν_L, arising from Yukawa interactions with the Higgs field, and,
- optionally, a Majorana mass term coupling ν_R with its charge-conjugated field $(\nu_R)^c$.

Considering the second possibility, it's straightforward to verify that:

$$(f^c)_L = (f_R)^c \qquad\qquad (f^c)_R = (f_L)^c \tag{1.4}$$

and it follows that the operator $\bar{\nu}_R(\nu_R)^c = \bar{\nu}_R(\nu^c)_L$ is Lorentz-invariant.

Such a Majorana mass term is permitted only if a neutrino is completely uncharged, that is, if:

$$\nu_{L,R} = (\nu_{L,R})^c = (\nu^c)_{R,L} \tag{1.5}$$

so that a neutrino is its own antiparticle.

But it's impossible to define a nonzero lepton number for such a field, and so a Majorana mass results in non-conservation of lepton number.

A gauge invariant and renormalizable Majorana mass for the three light neutrinos $\nu_{\ell L}$ of the standard model is *disallowed* because the light neutrinos carry both weak hypercharge and weak isospin.[1] But our postulated sterile neutrinos $\nu_{\alpha R}$ are uncharged under all gauge fields of the Standard Model, and are therefore permitted a Majorana mass.

1.1.2 Most general mass term

A generic Lagrangian term including both sources of mass—and accounting for the fact that the fields ν_ℓ are not mass eigenstates—must take the following form given in [2] and elsewhere:

$$-\mathcal{L}_m = \underbrace{M_{D\alpha\ell}\,\bar{\nu}_{\alpha R}\,\nu_{\ell L}}_{\text{Dirac term}} + \frac{1}{2}\,\underbrace{M_{N\alpha\beta}\,\bar{\nu}_{\alpha R}\,(\nu_{\beta R})^c}_{\text{Majorana term}} + \text{h.c.} \tag{1.6}$$

Here the fields $\nu_{\alpha R}$ are right-handed sterile neutrinos of index α.

The complex matrix M_D and real symmetric matrix M_N hold the coefficients of the Dirac and Majorana terms respectively.

Alternatively, the mass term (1.6) may be written in the form:

$$-\mathcal{L}_m = \frac{1}{2}\,\overline{\boldsymbol{\nu}^c}\,M_\nu\,\boldsymbol{\nu} + \text{h.c.} \qquad \text{where} \qquad \boldsymbol{\nu} = \begin{bmatrix}\boldsymbol{\nu}_L \\ \boldsymbol{\nu}_R^c\end{bmatrix} \qquad \overline{\boldsymbol{\nu}^c} = \begin{bmatrix}\overline{\boldsymbol{\nu}_L^c} & \bar{\boldsymbol{\nu}}_R\end{bmatrix} \qquad M_\nu = \begin{bmatrix}0 & M_D^T \\ M_D & M_N\end{bmatrix} \tag{1.7}$$

The top left corner of the matrix is zero because the left-handed fields are charged under the Standard Model gauge group and cannot have a renormalizable Majorana mass.

Since our mass matrix M_ν is manifestly non-diagonal, we must diagonalize it to find the fields of definite mass, which will in general be mixtures of the chiral fields. We observe that M_ν is symmetric, even though M_D is not.

1.1.3 Dirac neutrinos

If we take $M_N = 0$, and assume that there are exactly three sterile neutrino species $\nu_{\ell R}$, we obtain purely Dirac neutrinos. The attraction of this model is that it puts the neutrinos on the same footing as all other fermion fields in the Standard Model, with matching left and right chiral fields.

On the other hand, we may only introduce the new fields ν_R without also introducing the coupling M_N if we impose lepton number as a new global symmetry.

The Dirac mass matrix M_D may be diagonalized to $M_{\text{Dirac}} \equiv V_R^\dagger\,M_D\,V_L$ using a biunitary transformation, yielding a mass term of form:

$$-\mathcal{L}_m = \bar{\boldsymbol{\nu}}_{\text{Dirac}}\,M_{\text{Dirac}}\,\boldsymbol{\nu}_{\text{Dirac}} \tag{1.8}$$

[1] If we drop the requirement for renormalizability, and consider effective field theories, the dimension 5 Weinberg operator could give the light neutrinos a Majorana mass. We do not pursue this possibility here.

where the Dirac neutrino fields $\nu_{\text{Dirac}\,k}$ of definite mass are linear combinations of the chiral fields $\nu_{\ell L}$ and $\nu_{\ell R}$:

$$\boldsymbol{\nu}_{\text{Dirac}} = V_L^\dagger\,\boldsymbol{\nu}_L + V_R^\dagger\,\boldsymbol{\nu}_R \tag{1.9}$$

and are *not* purely left-handed.

A disappointing feature of this model is that it leaves unexplained the difference in mass between the neutrinos and the other fermions of the Standard Model, since the only sources of neutrino mass are the Yukawa interactions.

1.1.4 Majorana neutrinos

In the more general case $M_N \neq 0$, we may diagonalize M_ν with a unitary matrix V, so that $M_{\text{Maj}} \equiv V^T M_\nu V$ is diagonal. Then, making use of the relationships:

$$\Psi^c = C\gamma^0\Psi^\star \qquad\qquad \overline{\Psi^c} = -\Psi^T C^\dagger \qquad\qquad \overline{\Psi^c}^\dagger = -C\Psi^\star \tag{1.10}$$

where C is the charge conjugation matrix, the completely general mass term (1.7) becomes:

$$\begin{aligned}
-2\mathcal{L}_m &= \overline{\boldsymbol{\nu}^c}\,V^\star\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} + \boldsymbol{\nu}^\dagger\,V\,M_{\text{Maj}}\,V^T\,\overline{\boldsymbol{\nu}^c}^\dagger \\
&= -\boldsymbol{\nu}^T\,V^\star\,C^\dagger\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} - \boldsymbol{\nu}^\dagger\,V\,M_{\text{Maj}}\,C\,V^T\,\boldsymbol{\nu}^\star \\
&= \overline{(V^\dagger\boldsymbol{\nu})^c}\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,(V^\dagger\boldsymbol{\nu})^c \\
&= \overline{(V^\dagger\boldsymbol{\nu})^c}\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,(V^\dagger\boldsymbol{\nu})^c + \underbrace{\overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,V^\dagger\boldsymbol{\nu} - \left(\overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,V^\dagger\boldsymbol{\nu}\right)^T}_{=0} \\
&= \overline{(V^\dagger\boldsymbol{\nu})^c}\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,(V^\dagger\boldsymbol{\nu})^c + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,V^\dagger\boldsymbol{\nu} - \boldsymbol{\nu}^T V^\star\,M_{\text{Maj}}\,\gamma^0\,V^T\boldsymbol{\nu}^\star \\
&= \overline{(V^\dagger\boldsymbol{\nu})^c}\,M_{\text{Maj}}\,V^\dagger\,\boldsymbol{\nu} + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,(V^\dagger\boldsymbol{\nu})^c + \overline{V^\dagger\boldsymbol{\nu}}\,M_{\text{Maj}}\,V^\dagger\boldsymbol{\nu} + \overline{(V^\dagger\boldsymbol{\nu})^c}\,M_{\text{Maj}}\,(V^\dagger\boldsymbol{\nu})^c \\
&= \overline{\left[V^\dagger\boldsymbol{\nu} + (V^\dagger\boldsymbol{\nu})^c\right]}\,M_{\text{Maj}}\left[V^\dagger\boldsymbol{\nu} + (V^\dagger\boldsymbol{\nu})^c\right]
\end{aligned} \tag{1.11}$$

If we now define the Majorana fields $\boldsymbol{\nu}_{\text{Maj}} = \boldsymbol{\nu}^c_{\text{Maj}}$ of definite mass by:

$$\boldsymbol{\nu}_{\text{Maj}} \equiv V^\dagger\boldsymbol{\nu} + (V^\dagger\boldsymbol{\nu})^c \tag{1.12}$$

then we obtain a mass term in the form:

$$-\mathcal{L}_m = \frac{1}{2}\,\bar{\boldsymbol{\nu}}_{\text{Maj}}\,M_{\text{Maj}}\,\boldsymbol{\nu}_{\text{Maj}} \tag{1.13}$$

Again, the fields $\nu_{\text{Maj}\,i}$ are states of mixed chirality.

Let's consider three possibilities:

- If the Dirac masses vanish, $M_D = 0$, there are no terms mixing the left-handed fields $\nu_{\ell L}$ with the sterile neutrinos, and so the left-handed neutrinos remain massless, in contradiction to the experimental detection of neutrino flavor oscillations.
- On the other hand, we might have couplings of comparable magnitude, $M_N \sim M_D$. This would lead to more than three light neutrinos, each with a roughly comparable mix of left and right chirality.
- Finally, for $M_N \gg M_D$, we have a "see-saw", where the three small eigenvalues of M_ν are pushed downward, resulting in three very light Majorana neutrinos. The light neutrinos are mostly left-handed, and the remaining heavy neutrino fields are mostly right-handed. This is very consistent with known physics, and explains the anomalously tiny neutrino masses.

So if there are sterile neutrinos with Majorana-type mass terms, then the see-saw mechanism provides an elegant explanation for why neutrino masses are so much smaller than the masses of other elementary fermions, and this possibility is of great interest to BSM model builders. Sadly, neutrino oscillation experiments are incapable of distinguishing between Dirac and Majorana masses, so we do not discuss this fascinating question further.[2]

[2] Experiments to measure neutrino-less double beta decay are the most sensitive probe of whether neutrino masses are of Dirac or Majorana form.

1.1.5 Neutrino oscillations

In our discussion of neutrino oscillations, we'll mostly ignore matter effects, except for a brief description in section 1.2.7. The interaction of propagating neutrinos with matter is very important in the case of solar neutrinos—since the density of matter in the interior of the Sun is very high—but is a subdominant effect for the atmospheric neutrinos and accelerator experiments that are most relevant to this work.

Neutrino oscillations arise when the Hamiltonian for a free neutrino is not diagonal in the basis of fields $L_{\ell L}$ given in (1.1) above, that is, in the basis where the charged lepton fields e, μ, τ have definite mass. Off-diagonal terms of the Hamiltonian allow a free neutrino to transition between flavors in this basis.

These flavor transitions are "physical" because processes which produce neutrinos or detect neutrino flavor always involve charged lepton mass eigenstates. For example: solar neutrinos are produced by processes involving electrons, and atmospheric neutrinos are produced in processes involving electrons or muons. Similarly, a neutrino is observed when it interacts weakly with matter in the detector. For atmospheric and accelerator neutrinos, this is usually a charged current interaction, which couples states of definite flavor.

In the Standard Model, oscillations arise when we introduce a mass term with a mass matrix M which is non-diagonal in this basis, and couples neutrinos of different flavors.

1.1.6 Mass and interaction eigenstates

Whatever the source of neutrino mass, we denote the three neutrino *mass eigenstates* obtained by diagonalizing M as ν_1, ν_2, and ν_3, and the corresponding eigenvalues by m_1, m_2, and m_3. If these mass eigenstates were the same fields that appear in (1.1), that is, if M was already diagonal in the original basis, then any weak process producing neutrinos would produce the mass eigenstate corresponding to the charged lepton involved in the process, and there would be no neutrino oscillations.

For example, if ν_{1L} were the doublet-partner of e^-_L, then a process involving electrons would produce only neutrinos of type ν_1, which would freely propagate without any change of flavor. The famous deficit of solar neutrinos detected by the Homestake experiment [3] rules out this possibility.

Instead, the existence of flavor oscillations tells us that the neutrino fields which couple to the charged lepton mass eigenstates $\ell = e, \mu, \tau$ are *mixtures* of the mass eigenstates ν_i.

Generically, we indicate mass eigenstates with a latin subscript, for example, ν_i, ν_j, and weak interaction eigenstates with a greek subscript, for example, ν_α, ν_β.

1.1.7 Lepton mixing matrix

Weak interaction eigenstates $|\nu_\alpha\rangle$ are linear combinations of mass eigenstates $|\nu_i\rangle$.[3]

The lepton mixing matrix U, which is unitary, with $U^\dagger U = 1$, may be defined by the overlap:

$$U_{\alpha i} \equiv \langle \nu_\alpha | \nu_i \rangle \,, \qquad\qquad \alpha = e, \mu, \tau \quad ; \quad i = 1, 2, 3 \tag{1.14}$$

so that the following relationships (with implicit sums) follow quickly from the definition:

$$|\nu_i\rangle = |\nu_\alpha\rangle \langle \nu_\alpha | \nu_i \rangle = U_{\alpha i} \, |\nu_\alpha\rangle \,, \qquad |\nu_\alpha\rangle = |\nu_i\rangle \langle \nu_i | \nu_\alpha \rangle = U^\star_{\alpha i} \, |\nu_i\rangle = \left(U^\dagger\right)_{i\alpha} |\nu_i\rangle \tag{1.15}$$

For antineutrinos, the relationship corresponding to (1.15) is obtained by applying the charge conjugation operator $\mathcal{C}$ which satisfies $\mathcal{C}\psi_R = i\sigma^2 \psi^\star_L$ for two-component spinors:

$$|\bar\nu_\alpha\rangle = \mathcal{C} \, |\nu_\alpha\rangle = i\sigma^2 \, |\nu_\alpha\rangle^\star = i\sigma^2 \left(U^\star_{\alpha i} \, |\nu_i\rangle\right)^\star = U_{\alpha i} \, i\sigma^2 \, |\nu_i\rangle^\star = U_{\alpha i} \, \mathcal{C} \, |\nu_i\rangle = U_{\alpha i} \, |\bar\nu_i\rangle \tag{1.16}$$

where σ^2 is the second Pauli matrix. And so $U^\dagger$ is the mixing matrix for antineutrinos:

$$|\bar\nu_i\rangle = |\bar\nu_\alpha\rangle \langle \bar\nu_\alpha | \bar\nu_i \rangle = U^\star_{\alpha i} \, |\bar\nu_\alpha\rangle = \left(U^\dagger\right)_{i\alpha} |\bar\nu_\alpha\rangle \tag{1.17}$$

[3] We suppress the subscripts indicating chirality in this notation.

1.1.8 Parameterization of the lepton mixing matrix

With three neutrino flavors, the standard parameterization of the mixing matrix is given by [2], for example, as:

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{bmatrix} \begin{bmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}\, e^{-i\delta_{\mathrm{CP}}} \\ 0 & 1 & 0 \\ -\sin\theta_{13}\, e^{i\delta_{\mathrm{CP}}} & 0 & \cos\theta_{13} \end{bmatrix} \begin{bmatrix} \cos\theta_{21} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^{i\eta_1} & & \\ & e^{i\eta_2} & \\ & & 1 \end{bmatrix}$$

where the effects of the three mixing angles are separated into different matrices in the product.

The phases η_1, η_2 are nonzero only for Majorana neutrinos, and are not detectable in neutrino oscillation experiments,[4] so we will neglect them and use the simpler and more common form:

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{bmatrix} \begin{bmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}\, e^{-i\delta_{\mathrm{CP}}} \\ 0 & 1 & 0 \\ -\sin\theta_{13}\, e^{i\delta_{\mathrm{CP}}} & 0 & \cos\theta_{13} \end{bmatrix} \begin{bmatrix} \cos\theta_{21} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.18}$$

Here we have:

- three mixing angles θ_{ij}, along with
- a complex phase δ_{CP}.

If the phase δ_{CP} is zero, all elements of the mixing matrix are real-valued, and CP is conserved. But nonzero δ_{CP} results in non-conservation of CP.

If we narrow our attention to two-neutrino mixing, the 2×2 mixing matrix $U_{2\nu}$ may be parameterized as:

$$U_{2\nu} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{1.19}$$

where we have just one mixing angle θ between the two flavors, and no CP-violating phases.

Below, in 1.3, we will use a similar parameterization of NP effects, and explore the effect on neutrino oscillations. Right now it's time to consider the simpler case of neutrino oscillations in the absence of new physics.

1.2 Vacuum neutrino oscillations

We're interested in a neutrino produced at time $t = 0$ in an interaction eigenstate α which then propagates without interactions until arriving at a detector at some later time. Its evolution is completely determined by its kinetics.

But first we must consider the evolution of a mass eigenstate.

1.2.1 Evolution of mass eigenstates

As is standard, we treat a freely-propagating neutrino mass eigenstate of energy E_i as a stationary state in quantum mechanics, so that its time evolution is given by:

$$|\nu_i(x,t)\rangle = \mathcal{U}(t)\ |\nu_i(x,0)\rangle = \exp(-iH_{\mathrm{kinetic}}t)\ |\nu_i(x,0)\rangle = \exp(-iE_it)\ |\nu_i(x,0)\rangle \tag{1.20}$$

where $\mathcal{U}$ is the time evolution operator and H_{kinetic} is the Hamiltonian for a free neutrino.

If we take the state $|\nu_i(x,t)\rangle$ to represent a plane wave of definite momentum p_i then:

$$|\nu_i(x,t)\rangle = \exp(-iE_it + ip_ix)\ |\nu_i\rangle \tag{1.21}$$

Such a state does not represent a "physical" neutrino. Every experimentally observable neutrino is wavepacket superposition of momenta and of mass eigenstates. But for practical purposes the plane wave assumption is acceptable.

[4] See 1.2.3

1.2.2 Evolution of interaction eigenstates

On the other hand, the weak interaction eigenstates $|\nu_\alpha\rangle$ are *not* stationary states, and their time-dependence is more complicated. If we consider a neutrino of flavor α emitted at time $t = 0$, we may write:

$$|\nu_\alpha(x,t)\rangle = \mathcal{U}(t)\ |\nu_\alpha(x,0)\rangle = \exp(-iHt)\,|\nu_\alpha(x,0)\rangle \tag{1.22}$$

We must realize that $|\nu_\alpha(x,t)\rangle$ need not be proportional to $|\nu_\alpha\rangle$ at time $t > 0$, but rather may be a mix of interaction eigenstates, since H_{kinetic} is not diagonal in the interaction basis.

On the other hand, the neutrino at emission time $t = 0$ is a wavepacket superimposing plane waves of different momenta but the same interaction flavor α. Even though we're considering just one of the momentum modes here, we're entitled to assume that there's no dependence of flavor on x at the initial time. Therefore:

$$|\nu_\alpha(x,0)\rangle \propto |\nu_\alpha\rangle \tag{1.23}$$

and this allows us to use (1.15) at each position in space, so that:

$$|\nu_\alpha(x,0)\rangle = U^\star_{\alpha i}\,|\nu_i(x,0)\rangle = U^\star_{\alpha i}\ \exp(ip_i x)\,|\nu_i\rangle \tag{1.24}$$

where at the second step we used (1.21) with $t = 0$.

Inserting a full set of interaction flavors $1 = \sum_\beta |\nu_\beta\rangle\,\langle\nu_\beta|$ into (1.22), we find:

$$\begin{aligned}
|\nu_\alpha(x,t)\rangle &= \sum_\beta |\nu_\beta\rangle\,\langle\nu_\beta|\exp(-iHt)\,|\nu_\alpha(x,0)\rangle \\
&= \sum_\beta |\nu_\beta\rangle \underbrace{\Big(\sum_j U_{\beta j}\,\langle\nu_j|\Big)}_{\text{using (1.15)}} \exp(-iHt) \underbrace{\Big(\sum_i U^\star_{\alpha i}\ \exp(ip_i x)\,|\nu_i\rangle\Big)}_{\text{using (1.24)}} \\
&= \sum_{\beta,i,j} |\nu_\beta\rangle\,U_{\beta j}\,U^\star_{\alpha i}\ \exp(ip_i x)\,\langle\nu_j|\exp(-iHt)\,|\nu_i\rangle \\
&= \sum_{\beta,i,j} |\nu_\beta\rangle\,U_{\beta j}\,U^\star_{\alpha i}\ \exp(ip_i x)\exp(-iE_i t)\,\langle\nu_j|\nu_i\rangle \\
&= \sum_{\beta,i} U_{\beta i}\,U^\star_{\alpha i}\ \exp(-iE_i t + ip_i x)\,|\nu_\beta\rangle
\end{aligned} \tag{1.25}$$

This equation shows explicitly that a pure interaction flavor state ν_α evolves with time to produce a mix of interaction flavors, whenever the quantity $p_i x - E_i t$ varies between mass eigenstates ν_i. [5] On the other hand, if this quantity were the same for every ν_i, as it would be if the masses m_i were degenerate, then (1.25) would reduce to:

$$|\nu_\alpha(x,t)\rangle = \exp(-iEt + ipx)\,|\nu_\alpha\rangle \tag{1.26}$$

and, just as expected, we would have stationary states with no vacuum oscillations.

1.2.3 Transition probability

We may hit (1.25) from the left with the basis bra $\langle\nu_\beta|$ to obtain a transition amplitude we'll denote $\mathcal{A}(\alpha \to \beta)$:

$$\mathcal{A}(\alpha \to \beta) \equiv \langle\nu_\beta|\nu_\alpha(x,t)\rangle = \sum_i U_{\beta i}\,U^\star_{\alpha i}\ \exp(-iE_i t + ip_i x) \tag{1.27}$$

This is the amplitude for neutrinos produced in flavor α to be observed with flavor β at a later time t and position x.

[5] At first glance (1.25) looks like it contains the product $\sum_i U_{\beta i}\,U^\star_{\alpha i} = \delta_{\beta\alpha}$, but notice that the index i also occurs in the energy E_i and momentum p_i.

Since our neutrinos have tiny masses compared to their momenta, $m_i \ll p_i$, it's perfectly safe to treat them as highly relativistic particles, and:

$$E_i = \sqrt{p_i^2 + m_i^2} = p_i \sqrt{1 + \left(\frac{m_i}{p_i}\right)^2} \simeq p_i + \frac{m_i^2}{2p_i} \tag{1.28}$$

Furthermore, $x \simeq t$ and $p_i \simeq E$ for relativistic neutrinos propagating at close to the speed of light $c \equiv 1$, and so:

$$-E_i t + p_i x \simeq -\left(p_i + \frac{m_i^2}{2p_i}\right) t + p_i x = -\frac{m_i^2}{2p_i}\, t + p_i (x - t) \simeq -\frac{m_i^2}{2E}\, x \tag{1.29}$$

If we now consider a neutrino detector at $x = L$, we may write (1.27) as:

$$\mathcal{A}_L(\alpha \to \beta) \simeq \sum_i U_{\beta i}\, U_{\alpha i}^\star\, \exp\!\left(-i\, \frac{m_i^2}{2}\, \frac{L}{E}\right) \tag{1.30}$$

The transition probability is the square of this amplitude:

$$\begin{aligned} P_L(\alpha \to \beta) = \left|\mathcal{A}_L(\alpha \to \beta)\right|^2 &\simeq \left| \sum_i U_{\beta i}\, U_{\alpha i}^\star\, \exp\!\left(-i\, \frac{m_i^2}{2}\, \frac{L}{E}\right) \right|^2 \\ &= \left[\sum_i U_{\beta i}^\star\, U_{\alpha i}\, \exp\!\left(i\, \frac{m_i^2}{2}\, \frac{L}{E}\right) \right] \left[\sum_j U_{\beta j}\, U_{\alpha j}^\star\, \exp\!\left(-i\, \frac{m_j^2}{2}\, \frac{L}{E}\right) \right] \\ &= \sum_{i,j} U_{\beta i}^\star\, U_{\alpha i}\, U_{\beta j}\, U_{\alpha j}^\star\, \exp\!\left(\frac{i}{2}\, \frac{L}{E}\, \Delta m_{ij}^2\right) \end{aligned} \tag{1.31}$$

where at the last line we defined three squared-mass splittings $\Delta m_{ij}^2 \equiv m_i^2 - m_j^2$.

For convenience, we now also define:

$$X_{ij} \equiv \frac{1}{4}\, \frac{L}{E}\, \Delta m_{ij}^2 = -X_{ji} \qquad\qquad U_{\alpha\beta i} \equiv U_{\beta i}^\star\, U_{\alpha i} = U_{\beta\alpha i}^\star \tag{1.32}$$

Considering the case where the splittings vanish, $\Delta m_{ij}^2 = 0$, we see, again from the unitarity of the mixing matrix U, that $P_L(\alpha \to \beta) = \delta_{\alpha\beta}$ and there are no flavor oscillations. Similarly, and just as expected, the probability of a flavor transition vanishes at $L = 0$, or in the high-energy limit $E \to \infty$.

This is not the usual form in which the result is presented. Instead, to obtain the more standard form, we may split the sum (1.31) into terms with $i = j$ and terms with $i \neq j$:

$$\begin{aligned} P_L(\alpha \to \beta) &= \underbrace{\sum_{i=j} U_{\alpha\beta i}\, U_{\beta\alpha j}\, \exp(2iX_{ij})}_{\text{but notice } \Delta m_{ii}^2 = 0} + \underbrace{\sum_{i \neq j} U_{\alpha\beta i}\, U_{\beta\alpha j}\, \exp(2iX_{ij})}_{\text{also notice } \Delta m_{ij}^2 = -\Delta m_{ji}^2} \\ &= \sum_i |U_{\alpha\beta i}|^2 + \sum_{j>i} \left[U_{\alpha\beta i}\, U_{\beta\alpha j}\, \exp(2iX_{ij}) + U_{\alpha\beta j}\, U_{\beta\alpha i}\, \exp(2iX_{ji})\right] \\ &= \sum_i |U_{\alpha\beta i}|^2 + \sum_{j>i} \left[U_{\alpha\beta i}\, U_{\beta\alpha j}\, \exp(2iX_{ij}) + \text{c.c.}\right] \\ &= \underbrace{\sum_i |U_{\alpha\beta i}|^2}_{\text{average transition probability}} + \underbrace{2\,\text{Re} \sum_{j>i} U_{\alpha\beta i}\, U_{\beta\alpha j}\, \exp(2iX_{ij})}_{\text{sinusoidal } L\text{-dependence}} \end{aligned} \tag{1.33}$$

At $L = 0$, or in the case where the squared-mass splittings vanish, $\Delta m_{ij}^2 = 0$, the two terms in (1.33) must cancel when $\alpha \neq \beta$, though this is not obvious from staring at the equation.

Now, continuing from (1.33), we find:

$$\begin{aligned} P_L(\alpha \to \beta) &= \left(\sum_i U_{\alpha\beta i}\right)^2 - 2\,\mathrm{Re}\sum_{j>i} U_{\alpha\beta i}\,U_{\beta\alpha j} + 2\,\mathrm{Re}\sum_{j>i} U_{\alpha\beta i}\,U_{\beta\alpha j}\,\exp(2iX_{ij}) \\ &= \delta_{\alpha\beta} - 2\,\mathrm{Re}\sum_{j>i} U_{\alpha\beta i}\,U_{\beta\alpha j} \\ &\qquad + 2\,\mathrm{Re}\sum_{j>i}\Big[\mathrm{Re}\,(U_{\alpha\beta i}\,U_{\beta\alpha j}) + i\,\mathrm{Im}\,(U_{\alpha\beta i}\,U_{\beta\alpha j})\Big]\Big[\cos(2X_{ij}) + i\sin(2X_{ij})\Big] \qquad (1.34) \\ &= \underbrace{\delta_{\alpha\beta} - 4\sum_{j>i}\sin^2(X_{ij})\,\mathrm{Re}\,(U_{\alpha\beta i}\,U_{\beta\alpha j})}_{\text{CP conserving}} + \underbrace{2\sum_{j>i}\sin(2X_{ij})\,\mathrm{Im}\,(U_{\alpha\beta i}\,U_{\beta\alpha j})}_{\text{CP violating}} \qquad (1.35) \end{aligned}$$

Both oscillatory terms have the same underlying "normal mode" oscillation wavelengths:

$$\lambda_{\mathrm{osc},ij} = \frac{\pi}{X_{ij}/L} = \frac{4\pi E}{\Delta m^2{}_{ij}} \qquad (1.36)$$

With more than two flavors, these normal modes may interfere.

Notice that mixing matrix elements only occur as factors of the form $U^\star_{\beta i}\,U_{\alpha i}$, in which the Majorana phases η_i which we dropped back in (1.18) cancel, making no contribution to the transition probability.

1.2.4 Transition probability and CP violation

CP violation in (1.35) manifests as a difference in transition probabilities between the process $\nu_\alpha \to \nu_\beta$ and its CP-conjugate $\bar\nu_\alpha \to \bar\nu_\beta$. In light of (1.17), we see that the third term of (1.35) has opposite sign for antineutrinos, and is thus the source of CP violation.

This term is proportional to the Jarlskog invariant, J:

$$J \equiv \left|\mathrm{Im}\left(U^\star_{\alpha1}\,U_{\beta1}\,U_{\alpha2}\,U^\star_{\beta2}\right)\right| = \cos\theta_{12}\sin\theta_{12}\cos\theta_{23}\sin\theta_{23}\cos^2\theta_{13}\sin\theta_{13}\sin\delta_{\mathrm{CP}} \qquad (1.37)$$

which vanishes if any of the mixing angles $\theta_{ij} = 0$, or if the phase $\delta_{\mathrm{CP}} = 0$.

1.2.5 Transition probability with no CP violation

A final simplification is possible if we assume conservation of CP. In this case, the lepton mixing matrix U may be chosen to have all real elements, just by setting $\delta_{\mathrm{CP}} = 0$ in (1.18), and then (1.35) simplifies to:

$$P_L(\alpha \to \beta) = P_L(\bar\alpha \to \bar\beta) = \delta_{\alpha\beta} - 4\sum_{j>i} U_{\beta i}\,U_{\alpha i}\,U_{\beta j}\,U_{\alpha j}\,\sin^2\left(\frac{1}{4}\frac{L}{E}\,\Delta m^2_{ij}\right) \qquad (1.38)$$

In particular, for two neutrino flavors, where CP is always conserved, there's only one term with $i = 1, j = 2$, and we may plug in (1.19) to obtain the well-known formulas:

$$P_L(\alpha \to \beta) = \sin^2 2\theta\,\sin^2\left(\frac{1}{4}\frac{L}{E}\,\Delta m^2\right) \qquad \alpha \neq \beta \qquad (1.39)$$

$$P_L(\alpha \to \alpha) = 1 - \sin^2 2\theta\,\sin^2\left(\frac{1}{4}\frac{L}{E}\,\Delta m^2\right) \qquad (1.40)$$

The probability of observing a flavor transition is maximized by placing the detector at a distance:

$$L = \frac{2\pi E}{\Delta m^2}(2n+1), \qquad n = 0, 1, 2, \ldots \qquad (1.41)$$

from the point of emission of the neutrinos.

For experimental purposes the optimal choice of n is $n = 0$, since a real neutrino beam has a spectrum of energies, and the phase difference between the lowest-energy and highest-energy neutrinos accumulates with distance from the source.

1.2.6 Hamiltonian in the flavor basis

The Hamiltonian H_{kinetic} has so far has been implicit, but now let's write it out explicitly. In the mass basis, the Hamiltonian for kinetic vacuum oscillations is just:

$$H_{\text{kinetic}} = \begin{bmatrix} E_1 & & \\ & E_2 & \\ & & E_3 \end{bmatrix} = \begin{bmatrix} E_1 & & \\ & E_1 & \\ & & E_1 \end{bmatrix} + \begin{bmatrix} 0 & & \\ & \Delta E_{21} & \\ & & \Delta E_{31} \end{bmatrix} \tag{1.42}$$

It's acceptable to drop the uninteresting first term, as indicated, since it affects each mass eigenstate equally, and therefore makes no contribution to flavor oscillations.

But, starting from (1.28), we also have:

$$E_i \simeq p_i + \frac{m_i^2}{2p_i} \simeq E + \frac{m_i^2}{2E} \tag{1.43}$$

for neutrino mass eigenstates ν_i of mass m_i, momentum p_i, and energy $E_i = E$. And so:

$$H_{\text{kinetic}} \simeq \frac{1}{2E} \begin{bmatrix} 0 & & \\ & \Delta m_{21}^2 & \\ & & \Delta m_{31}^2 \end{bmatrix} \equiv \frac{1}{2E}\, \Delta M^2 \tag{1.44}$$

which leaves us with a Hamiltonian involving just two squared-mass splittings Δm_{21}^2 and Δm_{31}^2.

By convention (see [2], for example) we take all mixing angles within the first quadrant, $0 \leq \theta_{ij} < \frac{\pi}{2}$, and then choose the labels for our mass eigenstates such that the first of these splittings is smaller, $|\Delta m_{21}^2| < |\Delta m_{31}^2|$, and positive, $\Delta m_{21}^2 > 0$.

Neutrino oscillations are not sensitive to the absolute neutrino mass, which is not known, and depend only on the mass splittings. For example, observations of solar neutrinos are sensitive to Δm_{21}^2, and observations of atmospheric neutrinos are dominantly sensitive to Δm_{31}^2.

Indeed, as an aside, the ordering of the neutrino masses is still unknown, and there are two possibilities, which cannot be distinguished in current experiments. Given the convention just stated:

- the "normal" ordering is the case where $\Delta m_{31}^2 > 0$, and
- the "inverted" ordering is the case where $\Delta m_{31}^2 < 0$.

Finally, we transform our Hamiltonian to the weak interaction basis using the mixing matrix U, which—from now on—we'll call U^{kinetic} to avoid confusion:

$$H_{\text{kinetic}} = \frac{1}{2E}\, \Delta M^2 \qquad \underset{\text{mass basis to flavor basis}}{\longrightarrow} \qquad H_{\text{kinetic}} = \frac{1}{2E}\, U^{\text{kinetic}}\, \Delta M^2\, U^{\text{kinetic}\dagger} \tag{1.45}$$

The Hamiltonian we have just constructed is for neutrinos propagating in a vacuum. We have yet to account for the effect of matter.

1.2.7 Matter effects

Depending on the zenith angle, an atmospheric neutrino arriving at the detector might have passed all the way through the Earth. Even neutrinos in long-baseline accelerator experiments pass through the Earth's crust on their way to the detector. Therefore, our numerical calculations must take matter effects into account. We will not go into much detail here, since a more complete treatment may be found in [4] or [2].

Normal matter is composed of electrons, protons, and neutrons, described by their average densities n_e, n_p, and n_n, respectively. We assume the matter is electrically neutral so that $n_e = n_p$. Matter effects may be treated as an effective potential felt by a propagating neutrino.

The effective potential arises from two different elastic scattering processes of form $\nu\, f \to \nu\, f$:

- an electron neutrino may weakly interact with an electron via the charged current (CC), and
- a neutrino of any flavor may interact with an electron, proton, or neutron via the neutral current (NC).

The effective potentials V_{CC} and V_{NC} arising from these CC and NC interactions are:

$$V_{\text{CC}} = \sqrt{2}\, G_F\, n_e \qquad\qquad V_{\text{NC}} = \sum_{f=e,p,n} \sqrt{2}\, G_F\, n_f\, g_V^f \tag{1.46}$$

where G_F is the Fermi constant, and g_V^f is the vectorial coupling factor for fermion f. Thus, for neutrino flavor ℓ:

$$\begin{aligned} V_\ell &= V_{\text{CC}}\, \delta_{e\ell} + V_{\text{NC}} \\ &= \sqrt{2}\, G_F\, n_e \delta_{e\ell} + \sqrt{2} G_F \left(n_e g_V^e + n_p g_V^p + n_n g_V^n\right) \\ &= \sqrt{2}\, G_F\, n_e \delta_{e\ell} + \sqrt{2} G_F \left(n_e g_V^e + 2 n_p g_V^u + n_p g_V^d + n_n g_V^u + 2 n_n g_V^d\right) \\ &= \sqrt{2}\, G_F \left(n_e \delta_{e\ell} - \frac{1}{2}\, n_n\right) \end{aligned} \tag{1.47}$$

where the superscripts u and d refer to up and down quarks, and at the last line we made use of $n_e = n_p$ and found that the contributions from electrons and protons cancelled.

This effective potential leads to a contribution to H_{kinetic} which is diagonal in the flavor basis. But since the neutral current contribution V_{NC} is the same for every neutrino flavor ℓ, it does not contribute to neutrino oscillations, and we may ignore it.

Finally, our modified kinetic Hamiltonian is:

$$H_{\text{kinetic}} = \frac{1}{2E}\, U^{\text{kinetic}}\, \Delta M^2\, U^{\text{kinetic}\,\dagger} + V \qquad V = \begin{bmatrix} V_{\text{CC}} & & \\ & 0 & \\ & & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}\, G_F\, n_e & & \\ & 0 & \\ & & 0 \end{bmatrix} \tag{1.48}$$

Note that the matter effect has the opposite sign for antineutrinos propagating in regular matter.

Now let's see how the Hamiltonian changes when we add in our new physics.

1.3 Effect of new physics

In this work we're interested in flavor transitions that result not from the standard vacuum kinematic oscillations we considered in the previous section, but instead from some new interaction which also mixes neutrino flavors but arises from physics outside the framework of the Standard Model.

The effect of such an interaction is in principle detectable if it changes the energy dependence of the flavor oscillations. That is, the interaction may be distinguished from kinematic vacuum oscillations if it results in a departure from the linear dependence of λ_{osc} on E that we previously derived in (1.36).

NP effects falling into this category include:

- nonstandard interactions between neutrinos and matter [5],
- couplings with spacetime torsion fields [6],
- violations of Lorentz invariance [7] or of the equivalence principle [8], and
- violations of CPT symmetry [9].

Certain of the effects on the list were reviewed in [10]. In the past, these effects were proposed as competing explanations for neutrino oscillations, but today we may still consider them as subdominant effects.

We will not delve into the theory of any of this speculative new physics. Instead, we'll see a generic way to incorporate such effects into our Hamiltonian, along with a parametrization that's independent of the underlying physics.

1.3.1 Hamiltonian with new physics

Following [1], our generic Hamiltonian for NP effects will have a structure just like the Hamiltonian we've already seen in (1.45) for kinematic oscillations. However:

- there's no reason to suppose that generic NP interactions couple to the same interaction eigenstates as the weak interaction, and so we'll need a new mixing matrix, U^{NP}, and
- similarly, the resulting oscillations won't have the same $\Delta m^2/2E$ dependence we saw before, and so we'll introduce three quantities δ_i, playing a similar role to the squared masses m_i^2, and
- generalize the energy dependence to a generic power E^n.

Each NP effect exhibits a characteristic dependence on energy. The contribution of an effect with $n = -1$ cannot be distinguished in our experiments from the standard vacuum oscillations. In this work, we'll consider the cases $n = 0$, "vector-like" interactions, and $n = 1$, "tensor-like" interactions.

According to [1]:

- violation of the equivalence principle is tensor-like, with $n = 1$, while
- violation of CPT, or of Lorentz invariance is vector-like, with $n = 0$, and
- a coupling to a spacetime torsion field is also vector-like.

We may express an NP contribution to the Hamiltonian, H_{NP}, in the basis of weak interaction eigenstates, as:

$$H_{\text{NP}} = E^n \, U^{\text{NP}} \begin{bmatrix} \delta_1 & & \\ & \delta_2 & \\ & & \delta_3 \end{bmatrix} U^{\text{NP}\dagger} \tag{1.49}$$

Or, just like we did in (1.44), we may subtract off a quantity proportional to δ_1, and write H_{NP} as:

$$H_{\text{NP}} = E^n \, U^{\text{NP}} \begin{bmatrix} 0 & & \\ & \Delta\delta_{21} & \\ & & \Delta\delta_{31} \end{bmatrix} U^{\text{NP}\dagger} \equiv E^n \, U^{\text{NP}} \, \Delta\delta \, U^{\text{NP}\dagger} \tag{1.50}$$

Note that here we're considering just one source of new physics at a time. If there are multiple NP effects, we need multiple contributions of the above form. But in this work, we suppose that there's a single most-dominant NP effect, and that it's well-described by the equation above.

Finally, our full Hamiltonian H, incorporating kinematics, the NP contribution, and matter effects, could be expressed as:

$$H = H_{\text{kinetic}} + H_{\text{NP}} = \frac{1}{2E} \, U^{\text{kinetic}} \, \Delta M^2 \, U^{\text{kinetic}\dagger} + E^n \, U^{\text{NP}} \, \Delta\delta \, U^{\text{NP}\dagger} + V \tag{1.51}$$

We already have a parameterization of U^{kinetic}. Now we need to parameterize the matrix U^{NP}.

1.3.2 Parameterization of the NP mixing matrix

We will use uppercase latin indices, for example, ν_A, ν_B, to indicate eigenstates of the NP Hamiltonian. This lets us write down a definition for U^{NP}, precisely analogous to (1.14):

$$U^{\text{NP}}_{\alpha A} \equiv \langle \nu_\alpha | \nu_A \rangle \,, \qquad\qquad \alpha = e, \mu, \tau \quad ; \quad A = 1, 2, 3 \tag{1.52}$$

In section 1.1.8 we saw that a generic 3×3 unitary matrix may be parameterized by three mixing angles and three complex phases. That parameterization works just as well for U^{NP}.

Here we should keep all three phases around, since the two we dropped previously now represent relative phases between the kinetic mixing matrix and the NP mixing matrix.

We may thus write U^{NP} in the following form, in terms of:

- three mixing angles ξ_{ij}, and
- the "internal" phase η and relative phases α_1, α_2:

$$U^{\text{NP}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\xi_{23} & \sin\xi_{23} \\ 0 & -\sin\xi_{23} & \cos\xi_{23} \end{bmatrix} \begin{bmatrix} \cos\xi_{13} & 0 & \sin\xi_{13}\, e^{-i\eta} \\ 0 & 1 & 0 \\ -\sin\xi_{13}\, e^{i\eta} & 0 & \cos\xi_{13} \end{bmatrix} \begin{bmatrix} \cos\xi_{21} & \sin\xi_{12} & 0 \\ -\sin\xi_{12} & \cos\xi_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^{i\alpha_1} & & \\ & e^{i\alpha_2} & \\ & & 1 \end{bmatrix} \tag{1.53}$$

We'll also consider two-neutrino mixing, in which case the 2×2 mixing matrix $U_{\text{NP},2\nu}$ may be parameterized as:

$$U_{\text{NP},2\nu} = \begin{bmatrix} \cos\xi & \sin\xi\ e^{-i\eta} \\ -\sin\xi\ e^{i\eta} & \cos\xi \end{bmatrix} \tag{1.54}$$

where we have just one mixing angle ξ between the two flavors, along with a single relative phase η.

Actually, there's one more parameter we need to take into account.

1.3.3 NP effects and antineutrinos

Up to this point, we've had little to say about antineutrinos. In fact, everything we've said about neutrinos applies to their antiparticles. But there *is* one small thing we need to accommodate them into our framework.

In writing the NP Hamiltonian for an antineutrino, we will allow for the possibility that the NP effect is *CPT even* or *CPT odd*, that is, has the same or opposite sign for neutrinos and antineutrinos. [6]

To account for this, and recalling (1.17), we must modify our expression (1.51) for the full neutrino Hamiltonian H to obtain the antineutrino Hamiltonian $\bar{H}$:

$$\bar{H} = H_{\text{kinetic}} + \sigma\, H_{\text{NP}} = \frac{1}{2E}\, U^{\text{kinetic}\dagger}\, \Delta M^2\, U^{\text{kinetic}} + \sigma\, E^n\, U^{\text{NP}\dagger}\, \Delta\delta\, U^{\text{NP}} - V \tag{1.55}$$

where $\sigma = \pm1$, depending on whether the NP effect changes sign under conjugation of CPT.

1.3.4 NP oscillations for two neutrino flavors

We would now like to obtain a result analogous to (1.39) for the full Hamiltonian (1.51) in the two-flavor case with mixing matrices (1.19) and (1.54). We may solve this problem by simply diagonalizing H, to obtain a diagonal matrix D satisfying:

$$H = UDU^\dagger \qquad\qquad D = \begin{bmatrix} d_1 & \\ & d_2 \end{bmatrix} \tag{1.56}$$

where the eigenvalues d_1, d_2 are the energies of the stationary states of the full Hamiltonian.

The matrix U may be chosen to be unitary, since H is Hermitian. And so we may write:

$$U = \begin{bmatrix} \cos\Theta & \sin\Theta \\ -\sin\Theta & \cos\Theta \end{bmatrix} \tag{1.57}$$

for some mixing angle Θ, just as we saw in 1.1.8.

Following our usual practice, we may subtract away any constant term and express the Hamiltonian in terms of:

$$\Delta D \equiv \begin{bmatrix} 0 & \\ & \Delta d \end{bmatrix} \equiv \begin{bmatrix} 0 & \\ & d_2 - d_1 \end{bmatrix} \tag{1.58}$$

[6] Recall that the matter effects discussed in 1.2.7 are CPT odd.

Let's narrow our attention to the oscillation $\nu_\mu \to \nu_\tau$, where there are no matter effects and $V = 0$.

With an assist from Mathematica, we obtain:

$$\Delta d = \frac{\Delta m^2}{2E}\sqrt{1 + 4\left(E^{1+n}\frac{\Delta\delta}{\Delta m^2}\right)^2 + 4\left(E^{1+n}\frac{\Delta\delta}{\Delta m^2}\right)\Big(\cos 2\theta\cos 2\xi + \cos\eta\sin 2\theta\sin 2\xi\Big)} \tag{1.59}$$

This puts H in the same form as H_{kinetic} in the flavor basis, after making the replacement $\Delta M^2/2E \to \Delta D$ in (1.45). And so, simply by inspection, we may conclude that the mass splitting was "corrected" by a factor $\mathcal{R}$ given by:

$$\mathcal{R}(E) \equiv \sqrt{1 + R^2 + 2R\big(\cos 2\theta\cos 2\xi + \cos\eta\sin 2\theta\sin 2\xi\big)} \qquad R(E) \equiv 2E^{1+n}\frac{\Delta\delta}{\Delta m^2} \tag{1.60}$$

which is a result given in [1]. The corrected oscillation length is:

$$\lambda_{\text{osc}}^{\text{kinetic+NP}} = \frac{4\pi E}{\Delta m^2\,\mathcal{R}(E)} \tag{1.61}$$

We may take $\mathcal{R}$ to be small, at least in the limit $E \ll \Delta m^2/\Delta\delta$, and then:

$$\lambda_{\text{osc}}^{\text{kinetic+NP}} \simeq \frac{4\pi E}{\Delta m^2 + 2\Delta\delta\, E^{1+n}\left(\cos 2\theta\cos 2\xi + \cos\eta\sin 2\theta\sin 2\xi\right)} \tag{1.62}$$

which, in the cases we are considering, reduces to $\lambda_{\text{osc}}^{\text{kinetic}}$ in the low energy limit (or when $\Delta\delta$ vanishes).

Most importantly, oscillations no longer disappear in the high energy limit! Instead, taking $E \to \infty$:

- for $n = 0$, the oscillation wavelength approaches a constant limit, and
- for $n = 1$, the oscillation wavelength approaches zero.

Similarly, by inspecting the elements of U, we may obtain this formula given in [1] for the corrected mixing angle:

$$\sin^2 2\Theta(E) = \frac{1}{\mathcal{R}^2(E)}\Big(\sin^2 2\theta + R^2(E)\sin^2 2\xi + 2R(E)\sin 2\theta\sin 2\xi\cos\eta\Big) \tag{1.63}$$

Similar formulas exist in certain limits for three-flavor oscillations, some are given in [11], for example.

1.3.5 NP oscillations in general

Notice that the oscillation wavelength (1.61) is a complicated function of all the parameters of the Hamiltonian, even in the simplest case with just two flavors.

If H_{kinetic} and H_{NP} are simultaneously diagonalizable, that is, if they commute, then their contributions to the neutrino oscillation frequency combine linearly. But otherwise, the commutator $[H_{\text{kinetic}}, H_{\text{NP}}]$ encodes the first order nonlinear "interference" between the two contributions, since the time evolution operator is:

$$\mathcal{U}(t) = \exp(-itH_{\text{kinetic}} - itH_{\text{NP}}) \approx \exp(-itH_{\text{kinetic}})\exp(-itH_{\text{NP}})\exp\left(\frac{t^2}{2}\,[H_{\text{kinetic}}, H_{\text{NP}}]\right) \tag{1.64}$$

Therefore, to extract useful information about the oscillatory behavior, we must diagonalize H. However, with three neutrino flavors, H may only be diagonalized analytically in some limiting cases, and the expressions obtained are not particularly illuminating. Of course, oscillation probabilities may be found numerically.

It's now clear that the effect of introducing H_{NP} to the Hamiltonian was to:

- introduce a dependence on energy to the oscillation wavelength, with an effect which is negligible at low energy, but grows with increasing energy, and
- change the stationary states, by modifying the mixing angle, with an effect which also varies with energy.

If we perform experiments at a fixed energy E, these effects are effectively indistinguishable from a shift to the mixing angles and mass splittings of standard vacuum kinetic oscillations. But we can, in principle, tease out the NP effects by comparing data obtained with neutrinos of different energies.

1.3.6 Full parameterization

We now need to get used to thinking of H as a parameterized family of Hamiltonians:

$$H(\boldsymbol{\theta}) = H_{\text{kinetic}}(\boldsymbol{\theta}_{\text{kinetic}}) + H_{\text{NP}}(\boldsymbol{\theta}_{\text{NP}}) \tag{1.65}$$

Our parameter space is of $N+S$ dimensions, the product of an N–dimensional subspace of *NP parameters*, $\boldsymbol{\theta}_{\text{NP}}$, with an S–dimensional subspace of *kinetic parameters*, $\boldsymbol{\theta}_{\text{kinetic}}$. But the values of N and S depend on the whether we're considering two-flavor or three-flavor oscillations, and on whether we choose to neglect CP-violating phases.

With three flavors, the parameters $\boldsymbol{\theta}$ include:

- as kinetic parameters, the squared-mass splittings Δm^2_{ij}, kinetic mixing angles θ_{ij}, and strictly kinetic CP-violating phase δ_{CP},
- as NP parameters, the splittings $\Delta\delta_{AB}$, mixing angles ξ_{AB}, strictly NP "internal" phase η, and
- the relative complex phases α_i between H_{kinetic} and H_{NP} which as a convention we will assign to the NP sector.

Or, with just two flavors, we have:

- as kinetic parameters, the squared-mass splitting Δm^2 and mixing angle θ,
- as NP parameters, the strength of the NP effect, δ, and the mixing angle ξ, and
- the relative phase η.

Notice that, with three neutrino flavors:

- if there are no new physics, we have exactly one complex phase from the kinetic sector, and
- similarly, if neutrinos were massless, we would have exactly one phase from the NP sector, but
- with both massive neutrinos and new physics, we have *four* complex phases: δ_{CP}, η, α_1, and α_2.

In the two-flavor case, we have no CP-violating phases unless both sectors are present—that is, massive neutrinos with new physics–and then we have one relative phase η.

Parameter ranges are not uniquely defined. Due to symmetries of the Hamiltonian, we may sometimes trade a wider range of one parameter for a narrower range of another parameter.

This gets extremely complicated with three neutrino flavors. But in this work, we have not done any computations with the full set of parameters describing three-flavor mixing, so we don't need to discuss the parameter ranges of that description here. Some discussion may be found in Appendix B of [12].

With two neutrino flavors, the parameterization is simpler, and the parameter ranges are easier to nail down.

1.3.7 Two-flavor limit

We may obtain a limit with two-flavor oscillations simply by setting one of the mass splittings and two of the mixing angles to zero. Thus, there are actually two different useful two-flavor limits, but the one we're interested in is the limit which describes muon neutrino disappearance in atmospheric neutrino data.

The surviving parameters are:

- in the kinetic sector, $\theta \equiv \theta_{23}$ and $\Delta m^2 \equiv \Delta m^2_{31}$, along with,
- in the NP sector, $\xi \equiv \xi_{23}$ and $\Delta\delta \equiv \Delta\delta_{31}$, and
- the relative phase η.

The parameter ranges are given in [1]. For arbitrary η, they are:

Table 1.1: Parameter ranges for two-flavor oscillations

Parameter	Range	Alternate representation
Mass splitting	$\Delta m^2 \geq 0$	
NP splitting	$\Delta\delta \geq 0$	
Kinetic mixing angle	$0 \leq \theta \leq \frac{\pi}{2}$	$0 \leq \sin^2\theta \leq 1$
NP mixing angle	$0 \leq \xi \leq \frac{\pi}{4}$	$0 \leq \sin^2 2\xi \leq 1$
Relative phase	$0 \leq \eta \leq \pi$	

But if we limit our consideration to the subspace where the relative phase η is real, that is, $\eta \in \{0, \pi\}$, and CP is conserved, then we can drop η altogether if we let ξ range over negative values.

Table 1.2: Parameter ranges for two-flavor oscillations with CP conservation

Parameter	Range	Alternate representation
Mass splitting	$\Delta m^2 \geq 0$	
NP splitting	$\Delta\delta \geq 0$	
Kinetic mixing angle	$0 \leq \theta \leq \frac{\pi}{2}$	$0 \leq \sin^2\theta \leq 1$
NP mixing angle	$-\frac{\pi}{4} \leq \xi \leq \frac{\pi}{4}$	$-1 \leq \sin^2 2\xi \leq 1$

Results obtained assuming the two-flavor limit are documented in section 3.2.

A first step in generalizing from the two-flavor limit to incorporate the third neutrino flavor is the hierarchical approximation.

1.3.8 Hierarchical approximation

With three neutrino flavors, but without new physics, the *hierarchical approximation* described in [2] is the limit of vanishing Δm^2_{21}. In this limit, the ν_e appearance signal is *effectively* controlled by θ_{13} and the dependence on θ_{12} and δ_{CP} drops out, leaving Δm^2_{31}, θ_{23}, and θ_{13} as the relevant kinetic parameters. This is a useful limit for modelling long-baseline accelerator experiments, where the observables are ν_μ disappearance and ν_e appearance.

When adding new physics into the mix, we must be very careful in our selection of parameters, and in the interpretation we place on the results, since there is a risk of assigning effects to the NP sector which are actually explainable by unfitted parameters of the kinetic sector of the full model. The choice of NP parameters must be consistent with the choice of kinetic parameters.

In this case, since we excluded Δm^2_{21} and θ_{12} from the fit, we also exclude $\Delta\delta_{12}$ and ξ_{12}.

For simplicity, and since the effect of θ_{13} is already subdominant in our data, we will also exclude ξ_{13} from our fits, though in principle it could be included.

Furthermore, with Δm^2_{21}, $\Delta\delta_{21}$, and ξ_{13} set to zero, and following the logic outlined above in 1.3.6, we have only one complex phase, which we take to be α_2. From this point of view, the hierarchical approximation is still effectively a two-flavor limit.

A potential problem with this parameterization is that CP violation from the kinetic sector—for which some evidence exists in the appearance data—might be assigned to the relative phase between the kinetic and NP sectors.

If we further take θ_{13} to vanish, then we recover the two-flavor limit described in the previous section 1.3.7.

Incidentally, in (1.51) and (1.55) it can be seen that the effect of propagation in matter takes a similar form to a CPT odd NP effect with $n = 0$, and so it's important that matter be taken properly into account in our calculations, so as to avoid assigning its effect to the NP sector.

In section 3.3 we present the results of a number of computations performed under this hierarchical approximation. That is, instead of fitting all the parameters describing three-flavor oscillations, we excluded the parameters which were not well-constrained by the long-baseline dataset used in the fit, and fitted only the remaining parameters.

Chapter 2 || Statistical method

Our parameterized Hamiltonian $H(\boldsymbol{\theta})$ corresponds to a parameterized family of models we use to calculate probabilities which may be compared to the distribution of real data from neutrino oscillation experiments.

A given model—that is, a given assignment $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ of parameter values—determines a probability distribution of neutrino detection events. But, because of the complexity of the physics involved, it's not possible to express the distribution as a density function in closed form. Instead, it's necessary to use a numerical algorithm to estimate the discretized (binned) probability distribution.

Our goal in this work is to infer bounds on the parameters $\boldsymbol{\theta}_{\text{NP}}$ of H_{NP}. The approach we take, which will be rigorously justified below, involves:

- sampling the subspace of NP parameters on a grid, and, at each grid point
- using an optimization algorithm to find the model which best fits the experimental data within the subspace of kinetic parameters, and finally
- comparing the goodness of fit of the best model at each grid point with the best fit model across the whole grid, to determine an excluded region of the NP parameter space.

Let's see how we can make these statements precise.

2.1 Constructing confidence regions

Our ultimate task is parameter estimation, but to construct our confidence regions, we're going to make use of a statistic that's very often encountered in the context of tests for goodness of fit. The general framework for the methodology we adopt is presented in chapter 40 of [13], and a fascinating discussion contrasting the properties of different statistics for evaluation of goodness of fit may be found in [14].

The goodness of fit for any model is quantified by a χ^2 statistic, [1] and so a "best" fit is a model which minimizes χ^2 over some parameter space. Furthermore, the difference $\Delta\chi^2$ of χ^2 values between two different models may be used to precisely quantify how much "better" the first fit is compared to the second.

In principle, it would be perfectly acceptable to take a χ^2 value for a single model, and use its p-value directly to decide whether to exclude the corresponding point in parameter space. But our goal is parameter estimation. We're not interested in testing the overall fit of our parameterized model; rather, we're interested in excluding certain parameter assignments within the parameterized family.

Therefore, we always compare the χ^2 statistic for a model with the smallest value $\chi^2_{\text{best}} = \min(\chi^2)$ belonging to the best-fitting model in the given parameter space. This method does *not* attempt to decide whether there exists a model within the parameterized family which accurately captures all the real physics, and lends flexibility by accommodating the possibility that our family of models leaves some of the underlying physics unmodelled.

[1] Note that our χ^2 isn't just a sum of squares, but something more sophisticated, derived from the theory of maximum likelihood estimation, as we're about to see.

Calculating a χ^2 statistic is a rather complicated and computationally expensive process.

2.1.1 Simulation and binning

The experimental data is made available in binned form, with events binned by physical variables such as energy or zenith angle. Thus, the observed variables are the counts n_i of events falling within each bin i.

These observations must be compared to the expectations $\mu_i(\boldsymbol{\theta}_0)$ for each bin, predicted by a model with a given parameter assignment $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. These predictions are obtained by numerical integration of the flux and cross section, taking into account experimental efficiencies and resolution.

The first step of our data processing algorithm takes the n_i and $\mu_i(\boldsymbol{\theta}_0)$ together and produces a statistic $\chi^2(\boldsymbol{\theta}_0)$ representing the goodness of fit for the model at the point $\boldsymbol{\theta}_0$ in the full $N+S$–dimensional parameter space.

2.1.2 Goodness of fit for binned data

Each bin is modelled as an independent Poisson variable N_i with parameter μ_i as defined above:

$$N_i \sim \text{Poiss}(\mu_i) \tag{2.1}$$

If the number of events in the bin is large enough, a Gaussian approximation is acceptable:

$$N_i \sim \mathcal{N}(\mu_i, \mu_i) \tag{2.2}$$

For B bins with theoretical expectations $\boldsymbol{\mu}(\boldsymbol{\theta}_0) = (\mu_1, \ldots, \mu_B)$, the joint density function of the independent variables N_i is:

$$f_{\boldsymbol{\mu}}(\boldsymbol{n}) = \prod_{i=1}^{B} f_{\mu_i}(n_i) \tag{2.3}$$

Now, given binned experimental data $\boldsymbol{n} = (n_1, \ldots, n_B)$, this may be viewed as a likelihood function $L_m(\boldsymbol{\theta}_0) = f_{\boldsymbol{\mu}}(\boldsymbol{n})$ in the framework of maximum likelihood estimation.

This model will be evaluated by comparing it to a competing "saturated" model[2] in which we discard the model predictions μ_i and estimate the parameters of $f_{\boldsymbol{\mu}}$ directly from the experimental data, that is, we set $\mu_i = \hat{\mu}_i = n_i$, the usual maximum likelihood estimate of the Poisson parameter. Thus, we obtain a baseline likelihood $L_s(\hat{\boldsymbol{\mu}}) = f_{\boldsymbol{n}}(\boldsymbol{n})$. The saturated model should be considered a member of a parameterized family with B free parameters $\hat{\mu}_i$.

This is almost[3] the textbook setup for a likelihood ratio test. The likelihood ratio in question is:

$$\lambda(\boldsymbol{\theta}_0) \equiv \frac{L_m(\boldsymbol{\theta}_0)}{L_s(\hat{\boldsymbol{\mu}})} = \frac{f_{\boldsymbol{\mu}}(\boldsymbol{n})}{f_{\boldsymbol{n}}(\boldsymbol{n})} \tag{2.4}$$

Given a model with theoretical expectations $\boldsymbol{\mu}(\boldsymbol{\theta}_0)$, the ratio $\lambda(\boldsymbol{\theta}_0)$ compares the likelihood of the observations in the given model to their likelihood in the saturated model. A good fit is one in which λ is not much smaller than unity.

Finally, by Wilk's theorem,the following statistic has a χ^2 distribution with B degrees of freedom:

$$\chi^2(\boldsymbol{\theta}_0) \equiv -2 \log \lambda(\boldsymbol{\theta}_0) = -2 \log \frac{f_{\boldsymbol{\mu}}(\boldsymbol{n})}{f_{\boldsymbol{n}}(\boldsymbol{n})} \tag{2.5}$$

[2] See section 40.2.2.1 of [13].

[3] In the usual setup for a likelihood ratio test, our likelihood ratio would be $L_m(\hat{\boldsymbol{\theta}})/L_s(\hat{\boldsymbol{\mu}})$ and we would have $B - P$ degrees of freedom, since we would be estimating the P parameters θ_i from the data. But here we did *not yet* estimate any θ_i from data, they are fixed numbers baked into a model with effectively zero parameters.

For the case (2.1) of Poisson-distributed bins:

$$f_{\boldsymbol{\mu}}(\boldsymbol{n}) = \prod_{i=1}^{B} \frac{\mu_i^{n_i}}{n_i!} \exp(-\mu_i) \qquad \Rightarrow \qquad \lambda_{\text{Poiss}}(\boldsymbol{\theta}_0) = \frac{f_{\boldsymbol{\mu}}(\boldsymbol{n})}{f_{\boldsymbol{n}}(\boldsymbol{n})} = \prod_{i=1}^{B} \left(\frac{\mu_i}{n_i}\right)^{n_i} \exp(n_i - \mu_i)$$

And now, taking the log, we obtain equation (40.16) of [13]:

$$\chi^2_{\text{Poiss}}(\boldsymbol{\theta}_0) = -2 \log \lambda_{\text{Poiss}}(\boldsymbol{\theta}_0) = 2 \sum_{i=1}^{B} \left(n_i \log \frac{n_i}{\mu_i} - n_i + \mu_i \right) \tag{2.6}$$

In the Gaussian limit (2.2), we have something which looks initially more complicated, but is actually simpler:

$$f_{\boldsymbol{\mu}}(\boldsymbol{n}) = (2\pi)^{-\frac{B}{2}} \prod_{i=1}^{B} \frac{1}{\sigma_i} \exp\left(-\frac{1}{2} \frac{(n_i - \mu_i)^2}{\sigma_i^2} \right) \quad \Rightarrow \quad \lambda_{\mathcal{N}}(\boldsymbol{\theta}_0) = \frac{f_{\boldsymbol{\mu}}(\boldsymbol{n})}{f_{\boldsymbol{n}}(\boldsymbol{n})} = \prod_{i=1}^{B} \frac{\sqrt{n_i}}{\sigma_i} \exp\left(-\frac{1}{2} \frac{(n_i - \mu_i)^2}{\sigma_i^2} \right)$$

where σ_i^2 is the variance of N_i. We take the log and obtain:

$$\chi^2_{\mathcal{N}}(\boldsymbol{\theta}_0) = -2 \log \lambda_{\mathcal{N}}(\boldsymbol{\theta}_0) = \sum_{i=1}^{B} \left(\frac{(n_i - \mu_i)^2}{\sigma_i^2} - \log \frac{n_i}{\sigma_i^2} \right) \tag{2.7}$$

Recalling (2.2), we could now set $\sigma_i^2 = \mu_i$ and use this statistic directly. [4]

But instead, we replace σ_i^2 with its estimator $\hat{\sigma}_i^2 = \hat{\mu}_i = n_i$, leading to the simpler formula:

$$\chi^2_{\mathcal{N}}(\boldsymbol{\theta}_0) = \sum_{i=1}^{B} \frac{(n_i - \mu_i)^2}{n_i} \tag{2.8}$$

which is what [14] calls a "modified" or "Neyman's" χ^2, and is asymptotically equivalent when all n_i are sufficient. [5]

As mentioned above, we don't use this χ^2 statistic directly when constructing confidence regions.

2.1.3 Comparing models for goodness of fit

In this work, we must compare the goodness of fit for models at different points in parameter space.

This may be done by comparing their χ^2 statistics in a very trivial way since, if $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the parameter values for two different models, then:

$$\chi^2(\boldsymbol{\theta}_2) - \chi^2(\boldsymbol{\theta}_1) = -2 \log \frac{\lambda(\boldsymbol{\theta}_2)}{\lambda(\boldsymbol{\theta}_1)} = -2 \log \frac{L_m(\boldsymbol{\theta}_2)/L_s(\hat{\boldsymbol{\mu}})}{L_m(\boldsymbol{\theta}_1)/L_s(\hat{\boldsymbol{\mu}})} = -2 \log \frac{L_m(\boldsymbol{\theta}_2)}{L_m(\boldsymbol{\theta}_1)} \tag{2.9}$$

which is itself a negative log likelihood ratio.

Conveniently, the dependence on the saturated model has disappeared. And this trick generalizes to any case where two χ^2 values were computed from likelihood ratios with the same denominator.

Let's see how we can generalize this idea even further, to parameter subspaces.

2.1.4 Profile likelihood and nuisance parameters

Given some likelihood $L(\boldsymbol{\theta})$ defined on a parameter space $\boldsymbol{\theta}$, we may segregate the parameter space into two subspaces so that $\boldsymbol{\theta} = (\boldsymbol{\phi}, \boldsymbol{\nu})$ and, of course:

$$\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \max_{\boldsymbol{\phi}} \left[\max_{\boldsymbol{\nu}} L(\boldsymbol{\phi}, \boldsymbol{\nu}) \right] \tag{2.10}$$

[4] This formula is much simpler than the formula one would obtain using the MLE for a Gaussian distribution to estimate the μ_i in (2.4).

[5] Note that a textbook χ^2 statistic is usually given for a multinomial distribution, and has one less degree of freedom than the χ^2 statistic we just derived for our Poisson-distributed data.

This allows us to define the *profile likelihood*:[6]

$$L_{\text{profile}}(\boldsymbol{\phi}) \equiv \max_{\boldsymbol{\nu}} L(\boldsymbol{\phi}, \boldsymbol{\nu}) = L(\boldsymbol{\phi}, \hat{\boldsymbol{\nu}}(\boldsymbol{\phi})) \qquad \text{where} \qquad \hat{\boldsymbol{\nu}}(\boldsymbol{\phi}) \equiv \arg\max_{\boldsymbol{\nu}} L(\boldsymbol{\phi}, \boldsymbol{\nu}) \tag{2.11}$$

so that the maximum likelihood estimators of $\boldsymbol{\phi}$ and $\boldsymbol{\nu}$ are:

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} L_{\text{profile}}(\boldsymbol{\phi}), \qquad \hat{\boldsymbol{\nu}} = \hat{\boldsymbol{\nu}}(\hat{\boldsymbol{\phi}}) \tag{2.12}$$

This is a useful trick if the $\boldsymbol{\phi}$ are the parameters we're really interested in estimating, and the $\boldsymbol{\nu}$ are *nuisance parameters* that we would like to "profile out". In our work, the $\boldsymbol{\phi}$ will be NP parameters, and the $\boldsymbol{\nu}$ are usually a combination of kinetic parameters, NP parameters, and variables that encode systematic uncertainties.

We're not interested in the estimators, but in the *profile likelihood ratio*:[7]

$$\lambda_{\text{profile}}(\boldsymbol{\phi}) \equiv \frac{L_{\text{profile}}(\boldsymbol{\phi})}{L_{\text{profile}}(\hat{\boldsymbol{\phi}})} = \frac{L(\boldsymbol{\phi}, \hat{\boldsymbol{\nu}}(\boldsymbol{\phi}))}{L(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}})} \tag{2.13}$$

A straightforward application of Wilk's theorem shows that:

$$\Delta\chi^2_{\text{profile}}(\boldsymbol{\phi}) \equiv -2\log\lambda_{\text{profile}}(\boldsymbol{\phi}) = 2\log L(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}) - 2\log L(\boldsymbol{\phi}, \hat{\boldsymbol{\nu}}(\boldsymbol{\phi})) \tag{2.14}$$

approaches a χ^2 distribution where the degrees of freedom are the same as the number of parameters in $\boldsymbol{\phi}$ when the sample size is very large.

This is a great result: it tells us that we can profile away the nuisance parameters simply by finding the best-fitting model within the nuisance parameter subspace, and still meaningfully compare models within the complementary subspace of parameters of interest.

2.1.5 Excluding models

Best of all, $\Delta\chi^2_{\text{profile}}$ is easy to calculate. Reversing the logic of (2.9), we see that:

$$\Delta\chi^2_{\text{profile}}(\boldsymbol{\phi}) = -2\log\frac{L_{\text{profile}}(\boldsymbol{\phi})}{L_{\text{profile}}(\hat{\boldsymbol{\phi}})} = -2\log\frac{\lambda(\boldsymbol{\phi}, \hat{\boldsymbol{\nu}}(\boldsymbol{\phi}))}{\lambda(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}})} = \chi^2(\boldsymbol{\phi}, \hat{\boldsymbol{\nu}}(\boldsymbol{\phi})) - \chi^2(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\nu}}) \tag{2.15}$$

Thus, to determine an excluded region of parameter space—for *any* selection of NP parameters—we simply:

1. place a grid $\boldsymbol{\phi}_{ij}$ over the selected parameter space, and then,
2. at each grid vertex (i, j), minimize $\chi^2(\boldsymbol{\phi}_{ij})$ over the parameters we're *not* interested in, and finally
3. evaluate $\Delta\chi^2_{\text{profile}}(\boldsymbol{\phi}_{ij})$ at each vertex (i, j) by subtracting the best (smallest) χ^2 on the whole grid from the value $\chi^2(\boldsymbol{\phi}_{ij})$ at the vertex.

We may use the p-value of $\Delta\chi^2_{\text{profile}}(\boldsymbol{\phi}_{ij})$ to determine if the vertex $\boldsymbol{\phi}_{ij}$ is excluded.

2.1.6 Systematic uncertainties

Systematic uncertainties—that is, errors which do not converge to zero with increasing sample size—may also be treated as nuisance parameters. Without loss of generality, we will take each of these m parameters ξ_i to be "centered" at zero. That is, we normalize the parameters so that our best prior estimate of ξ_i before fitting is $\hat{\xi}_i = 0$.

We must also incorporate our best estimate of the magnitude of each uncertainty, and for this we introduce a *regularization factor* (a Gaussian prior) for each parameter ξ_i, which modifies the likelihood function L.

[6] See section 40.2.2.2 of [13].
[7] See section 40.3.2.1 of [13].

The regularized likelihood L_{reg} is given by:

$$L_{\text{reg}}(\boldsymbol{\theta}, \boldsymbol{\xi}) = L(\boldsymbol{\theta}, \boldsymbol{\xi})\,(2\pi)^{-\frac{m}{2}} \prod_{i=1}^{m} \frac{1}{\sigma_i} \exp\left(-\frac{1}{2}\frac{\xi_i^2}{\sigma_i^2}\right) \tag{2.16}$$

where the σ_i encode our prior uncertainty about the values of ξ_i.

Now, after dropping constant terms, the regularized log likelihood is just:

$$\log L_{\text{reg}}(\boldsymbol{\theta}, \boldsymbol{\xi}) = \log L(\boldsymbol{\theta}, \boldsymbol{\xi}) - \frac{1}{2}\sum_{i=1}^{m} \frac{\xi_i^2}{\sigma_i^2} \tag{2.17}$$

and, making use of (2.5), we may obtain the statistic:

$$\chi^2(\boldsymbol{\theta}, \boldsymbol{\xi}) = -2\log \lambda_{\text{reg}}(\boldsymbol{\theta}, \boldsymbol{\xi}) = -2\log \lambda(\boldsymbol{\theta}, \boldsymbol{\xi}) + \sum_{i=1}^{m} \frac{\xi_i^2}{\sigma_i^2} \tag{2.18}$$

If we now assume a model with Gaussian-distributed bins, as above in (2.8), we obtain:

$$\chi^2(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sum_{j=1}^{B} \frac{1}{n_j}\Big[\mu_j(\boldsymbol{\theta}, \boldsymbol{\xi}) - n_j\Big]^2 + \sum_{i=1}^{m} \frac{\xi_i^2}{\sigma_i^2} \tag{2.19}$$

$$\simeq \sum_{j=1}^{B} \frac{1}{n_j}\left[\mu_j(\boldsymbol{\theta}, 0) + \sum_{i=1}^{m} \xi_i \frac{\partial \mu_j}{\partial \xi_i}(\boldsymbol{\theta}, 0) - n_j\right]^2 + \sum_{i=1}^{m} \frac{\xi_i^2}{\sigma_i^2} \tag{2.20}$$

Following the logic of section 2.1.4, we should minimize this χ^2 over the nuisance subspace $\boldsymbol{\xi}$.

But at a minimum of χ^2 we must have, for each ξ_i:

$$0 = \frac{\partial \chi^2(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial \xi_i} \simeq \sum_{j=1}^{B} \frac{2}{n_j}\frac{\partial \mu_j}{\partial \xi_i}\left[\mu_j + \sum_{k=1}^{m} \frac{\partial \mu_j}{\partial \xi_k}\,\xi_k - n_j\right] + \frac{2}{\sigma_i^2}\,\xi_i \tag{2.21}$$

and this is a system of linear equations for the ξ_i, and can be easily solved.

A difficulty here is obtaining the derivatives $\partial \mu_j / \partial \xi_i$, but if analytic expressions are not available, they may be estimated using numerical differentiation.

2.2 Optimization

In the previous section we encountered the need to find the best-fitting model over a subspace of the whole parameter space. In fact, this need arises in three scenarios:

1. when marginalizing over systematic uncertainties, as discussed above in 2.1.6, although, as we've just seen, this is a special case with an easy solution,
2. when, given P parameters $\boldsymbol{\phi}$, we find the best-fitting model over the remaining subspace of $N+S-P$ parameters $\boldsymbol{\nu}$, as discussed in 2.1.4, and
3. when we compare the best-fitting model for a given $\boldsymbol{\phi}$ to the best-fitting model over the entire $N+S$–dimensional parameter space, as discussed in 2.1.5.

Here we discuss how best fits were obtained in each scenario. Actually, we've already seen the solution used in the first scenario. For the second scenario, we must resort to the use of optimization algorithms. But for the third scenario, a brute force approach is possible, so let's review that option first.

2.2.1 Brute force approach

A brute-force approach to finding the best fit is always possible. For example, in a two-dimensional subspace, we might place a grid $\boldsymbol{\phi}_{ij}$ over the space, evaluate $\chi^2(\boldsymbol{\phi}_{ij})$ for the model at each grid vertex, and then simply select the smallest value. This approach is perfectly acceptable in two dimensions, but scales badly to higher-dimensional subspaces. Since, as mentioned above, the evaluation of χ^2 is computationally expensive, we quite quickly saturate our performance constraints.

Nevertheless, brute force *is* the approach we take for the third item on the list above. The reason it's appropriate in that case is that we are already evaluating χ^2 at every point on a grid when determining an excluded region in NP parameter space. Furthermore, in practice, $P = 2$ here, since we produce two-dimensional plots.

On the other hand, we do not use brute force to find the best-fitting model for a given $\boldsymbol{\phi}$ since our ultimate goal is to develop a methodology suitable for fits over the full set of standard and NP parameters, where $N+S-P$ is large enough that a brute force search is prohibitively computationally expensive. Instead, we must use an optimization algorithm to minimize χ^2 over the subspace.

2.2.2 Optimization algorithms

Our optimization problem is not a trivial one since:

- evaluations of the χ^2 function we're trying to minimize are very expensive,
- the χ^2 function itself does not vary monotonically in all directions as we move away from the global minimum, having oscillations in certain directions in some regions of the parameter space, and
- the parameter space has up to 14 dimensions.

In this section, we suppose that our task is to minimize an objective function $f(\boldsymbol{x})$ over values of the parameters $\boldsymbol{x} = (x_1 \dots x_P)$. In this work, f is almost always χ^2, and $\boldsymbol{x}$ represents the same parameters we called $\boldsymbol{\theta}$ in the previous section.

The optimization algorithms we consider belong to three families:

- *gradient descent*, relying on numerical differentiation to determine the slope of the objective function,
- *direct methods*, which do not require the use of a local gradient, and
- *population methods*, which are inspired by analogy to biological systems, and work by propagating successful traits between members of the population.

We now discuss the algorithms we used, which the author implemented in C, based on descriptions in the excellent reference [15].

2.2.3 Gradient descent

In the most basic gradient descent method, we choose a *learning rate* α, then pick an initial point $\boldsymbol{x}$ and iteratively move in the opposite direction of the gradient ∇f of f, according to:

$$\boldsymbol{x} \leftarrow \boldsymbol{x} - \alpha \nabla f(\boldsymbol{x})$$

Some variations on gradient descent also feature a decaying *momentum*–it's actually just a velocity—which helps prevent the algorithm taking many iterations to traverse a flat region. With momentum $\boldsymbol{v}$ we have:

$$\begin{aligned} \boldsymbol{x} &\leftarrow \boldsymbol{x} + \boldsymbol{v} \\ \boldsymbol{v} &\leftarrow \beta \boldsymbol{v} - \alpha \nabla f(\boldsymbol{x}) \end{aligned}$$

where the hyperparameter $1-\beta$ must be some sort of linear drag, the author supposes.

Gradient descent algorithms are very susceptible to getting stuck in local minima, but this problem may be partially alleviated by sampling many random initial points in the parameter space, and choosing the best result.

Our objective function has no expression in closed form, and so we can't differentiate it. This makes gradient descent a potentially-poor choice for our needs. Nevertheless, we may use numerical differentiation to obtain an estimate of the gradient at any given point.

We tested two gradient descent algorithms:

- gradient descent with *Nesterov momentum*[16], a slightly fancier version of what we just described, and
- *Adam*[17], or *adaptive moment estimation*, a sophisticated gradient descent algorithm that's very popular in the Machine Learning community.

Nesterov momentum evaluates the gradient not at the current position $\boldsymbol{x}$, but at a projected position $\boldsymbol{x} + \beta\boldsymbol{v}$. Thus, the rule to update the momentum becomes:

$$\boldsymbol{v} \leftarrow \beta\boldsymbol{v} - \alpha\nabla f(\boldsymbol{x} + \beta\boldsymbol{v})$$

We did have some success with the Nesterov algorithm as long as we started it with sufficiently-many random initial points. However, it was not as efficient as other methods considered below. This is hardly surprising, since numerical differentiation of an objective function like ours is not exactly a thing that's obviously going to work.

We had little success with Adam, and quickly discarded our implementation.

Ultimately we did not use gradient descent to obtain the results presented here.

2.2.4 Direct methods

More promising, for our particular objective function, are what [15] calls *direct methods*. A direct method uses only the objective function f and none of its derivatives.

The Nelder-Mead simplex method[18, 19] is a very famous direct method. A simplex is a generalization of a triangle or tetrahedron to arbitrary dimensions. In P dimensions, a simplex is characterized by its $P+1$ vertices, and is thus the simplest geometric figure with the given dimensionality.

The simplex method starts with an initial simplex $(\boldsymbol{x}^{(1)} \ldots \boldsymbol{x}^{(P)})$, along with hyperparameters α, β, and γ, and at each iteration determines the best and worst vertices $\boldsymbol{x}^{(\text{best})}$ and $\boldsymbol{x}^{(\text{worst})}$, computes the centroid $\bar{\boldsymbol{x}}$, and then either:

- *reflects* the worst vertex about the centroid:

$$\boldsymbol{x}^{(\text{worst})} \leftarrow (1+\alpha)\,\bar{\boldsymbol{x}} - \alpha\,\boldsymbol{x}^{(\text{worst})},$$

- reflects the worst vertex about the centroid and also *expands* the simplex:

$$\boldsymbol{x}^{(\text{worst})} \leftarrow (1+\alpha\beta)\,\bar{\boldsymbol{x}} - \alpha\beta\,\boldsymbol{x}^{(\text{worst})},$$

- *contracts* the simplex away from the worst vertex:

$$\boldsymbol{x}^{(\text{worst})} \leftarrow (1-\gamma)\,\bar{\boldsymbol{x}} + \gamma\,\boldsymbol{x}^{(\text{worst})},$$

- or *shrinks* the simplex by moving all vertices toward the best vertex:

$$\boldsymbol{x}^{(p)} \leftarrow \frac{1}{2}\left(\boldsymbol{x}^{(\text{best})} + \boldsymbol{x}^{(p)}\right).$$

A complete definition of the algorithm, including the criteria for choosing the action to take at each step, may be found in any of the references cited above.

We found that this method worked very well for our problem.

2.2.5 Population methods

A population method begins with a population of randomly-sampled individuals $\boldsymbol{x}^{(p)}$. Each iteration attempts to improve the average "fitness" of the population by moving individuals in directions that reduce the objective function, and—unlike other optimization methods—the individuals are allowed to share information so that successful "traits" propagate within the population.

Note that this is quite different to simply starting a descent algorithm from many randomly-chosen initial points. In a population method, the individuals cooperate to discover the optimal solution.

The population methods we tested were:

- *differential evolution* [20, 21], which works by recombining individuals in the population, retaining only those recombinant individuals with improved fitness compared to the previous generation, and
- *particle swarm* [22], which uses a sort of momentum, and each individual is accelerated toward the best position it has visited, and the best position found by the whole population.

Differential evolution updates a population member $\boldsymbol{x}^{(p)}$ by "crossing" it with three other individuals:

1. For three distinct randomly-selected individuals $\boldsymbol{x}^{(q)}$, $\boldsymbol{x}^{(r)}$, $\boldsymbol{x}^{(s)}$, let:
$$\boldsymbol{z} = \boldsymbol{x}^{(q)} + w\Big(\boldsymbol{x}^{(r)} - \boldsymbol{x}^{(s)}\Big)$$
where the weight w is a hyperparameter.
2. Construct a recombinant individual $\boldsymbol{x}'$ from $\boldsymbol{x}^{(p)}$ and $\boldsymbol{z}$. For each component j, make a random choice:
$$x'_j = \begin{cases} z_j & \text{if } r_j < p \\ x_j^{(p)} & \text{otherwise} \end{cases}$$
where the $0 \leq r_j \leq 1$ are random numbers, and the probability p is a hyperparameter.
3. If $f(\boldsymbol{x}') < f(\boldsymbol{x}^{(p)})$, set $\boldsymbol{x}^{(p)} \leftarrow \boldsymbol{x}'$. Otherwise, retain the previous individual.

There are many possible minor variations on this basic scheme.

Differential evolution is sometimes known in the particle physics community by the name "Diver", referring to a particular implementation in Fortran [23]. The author invested some time in testing this package and reviewing the implementation. Finding nothing special in its feature set, and considering the implementation rather bloated and over-complex, the author quickly implemented a generic differential evolution algorithm in a few lines of C, and, by luck rather than design, found that it slightly outperformed Diver configured with default settings, at least for our particular χ^2 objective function. We made no further use of the Diver package.

Our implementation has a number of tunable hyperparameters, some copied from [23]. We were unable to demonstrate that any of these knobs and switches improved performance over the most generic implementation of differential evolution. For some other objective function, these extra options would surely be useful.

Particle swarm is, if we squint, similar to gradient descent with momentum, except that there's no need to compute a local gradient. Instead, the direction of acceleration is determined by positions visited in the past. Each population member has its position $\boldsymbol{x}_p^{(i)}$ and momentum $\boldsymbol{v}_p^{(i)}$ updated according to:

$$\begin{aligned} \boldsymbol{x}^{(p)} &\leftarrow \boldsymbol{x}^{(p)} + \boldsymbol{v}^{(p)} \\ \boldsymbol{v}^{(p)} &\leftarrow w\boldsymbol{v}^{(p)} + c_1 r_1 \Big(\boldsymbol{x}_{\text{best}}^{(p)} - \boldsymbol{x}^{(p)}\Big) + c_2 r_2 \Big(\boldsymbol{x}_{\text{best}} - \boldsymbol{x}^{(p)}\Big) \end{aligned}$$

where the $0 \leq r_i \leq 1$ are random numbers, c_1, c_2, and w are hyperparameters, $\boldsymbol{x}_{\text{best}}^{(p)}$ is the best position visited by the individual, and $\boldsymbol{x}_{\text{best}}$ is the best position found so far by the whole population.

For our objective function, the particle swarm method was quite similar differential evolution when the two methods were compared by efficiency.

2.2.6 Hyperparameters

Each of the optimization algorithms we considered came with several "knobs", hyperparameters that in principle allow the algorithm itself to be tuned for better performance. These hyperparameters may be adjusted by hand, but a better approach is an automated search of the hyperparameter space, a practice that's routine at least in in the context of Machine Learning.

It naturally occurred to us then, that since we happened to have some nice optimization algorithms on hand, why not use the optimization algorithm itself to optimize its own hyperparameters? For this problem, the objective function is not the χ^2 function, but the number of evaluations of χ^2 that are required for the algorithm to converge and terminate.

Unfortunately, what we found was that the optimization algorithm did not usually converge when optimizing its own hyperparameters, or, when it did converge, it converged to what were obviously local minima. We attribute this failure to the inherent nondeterminism in the algorithms themselves—even the deterministic algorithms were started at randomly sampled initial points—combined with the tricky oscillatory behavior of our χ^2 function.

If this explanation is correct, then attempting to tune the hyperparameters by hand is even less likely to be successful, and, indeed, we were not successful when we tried.

2.2.7 Comparison of optimization algorithms

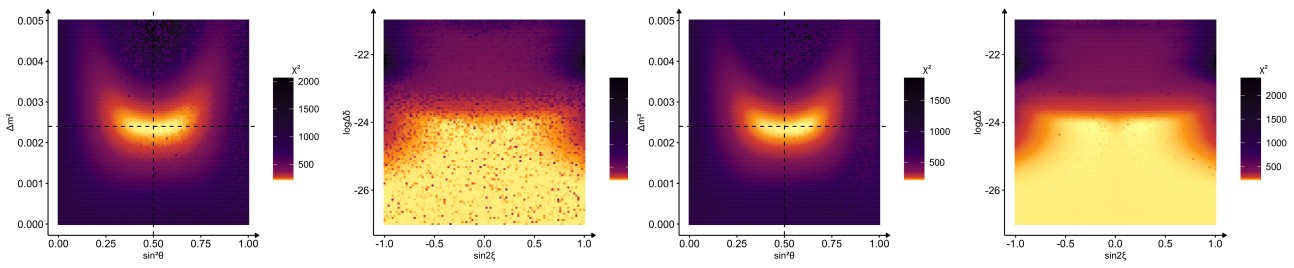
The most efficient optimization algorithm is the one which evaluates the objective function the fewest times on its way to convergence. Implicit in this definition is that the convergence criteria of the algorithms should be sufficiently similar for this comparison to be fair, and that they should be converging to the same global minimum. It's not very clear if these conditions are perfectly satisfied here.

A semi-qualitative method of evaluating an algorithm is to:

1. select two NP parameters of interest, and place a grid on the two-dimensional parameter space,
2. set up the algorithm to find the minimum of χ^2 over the remaining parameters, and then
3. by trial and error, determine a sufficiently-large population size (or, equivalently, a sufficiently-large sample size of random initial points) such that a visually-smooth heatmap is produced when χ^2 is plotted against the two parameters of interest, and
4. report how many evaluations of χ^2 were performed by the algorithm.

We found that non-smoothness of the heatmap produced in step 3 was an effective way to detect when an algorithm was failing to converge, getting lost in noise, or stuck in local minima. These failures tend to result in curves of discontinuity, or visible "speckling" in the heatmap, as seen, for example, in figure 2.1.

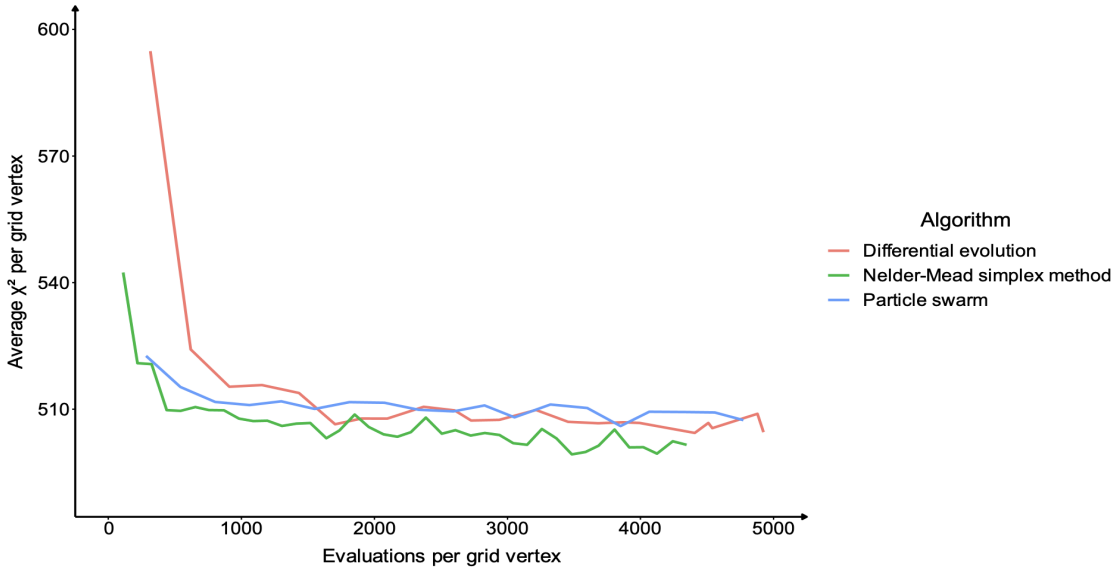
Figure 2.1: Heatmap of χ^2_{profile} values produced by particle swarm optimization with population sizes 5 (plots on left) and 20 (plots on right).



By this method we were able to narrow our selection of algorithms to the following three: the Nelder-Mead simplex method, differential evolution, and particle swarm. To decide between them, we'll need something a bit more quantitative.

Figure 2.2 plots the average (over a 2×2 grid) number of evaluations of the objective function for a model with four free parameters which we will meet in section 3.3, against the average minimized value of χ^2 obtained by the algorithm, while varying the population size or number of tries per grid point, for the three algorithms which were found to be competitive.

Figure 2.2: Comparison of optimization algorithm performance



We emphasize that this is far from a truly "fair" evaluation:

- we have not tuned the hyperparameters of each algorithm for best performance, and
- the convergence criteria for the population methods is by nature quite different to that used for the simplex method.

But our objective in this work is not a rigorous evaluation of optimization algorithms. We're merely trying to pick an algorithm that produces acceptable results, efficiently, and with a minimum of fuss. We didn't have time to spend weeks tuning hyperparameters.

The results we present in the next chapter were all obtained using the simplex method.

Chapter 3 || Results

Here we present the results of our computations. We began by fitting the parameters of the two-flavor limit from section 1.3.7 to data sensitive to the mass difference Δm^2_{31} and mixing angle θ_{23}, mainly as a way of validating our code, and providing a baseline for comparison. We then moved on to fitting the parameters of the hierarchical approximation described in section 1.3.8, by introducing θ_{13} into the fit.

3.1 Experimental datasets

The results in this section have been produced using data from two different sorts of experiment.

- Experiments with *atmospheric neutrinos*, measuring the disappearance of muon neutrinos. When cosmic rays interact with the atmosphere, pions are produced, and rapidly decay. Decaying pions produce muon neutrinos and muons. The muons themselves decay, producing more muon neutrinos, along with an equal number of electron neutrinos.

  Experiments of this sort are Super-Kamiokande (SK) and DeepCore. Neutrinos detected by SK have energies between 10^{-1} and 10^4 GeV.

  These experiments are most sensitive to the "atmospheric" mixing angle θ_{23}, that is, to the oscillation $\nu_\mu \to \nu_\tau$, but θ_{13} plays a subdominant role.
- *Long-baseline accelerator* experiments, measuring the disappearance of ν_μ from a muon neutrino beam, over distances of hundreds of kilometers. In a long-baseline experiment it's also possible to observe the appearance of electron neutrinos, due to the oscillation $\nu_\mu \to \nu_e$.

  Compared to atmospheric neutrinos, the muon neutrinos have relatively low energies in the realm $1-10$ GeV. Note that it's very difficult to observe appearance of tau neutrinos, since the energy of the beam is small compared to the τ mass.

  Experiments include KEK to Kamioka (K2K), Tokai to Kamioka (T2K), MINOS, and NOνA. These experiments are sensitive to θ_{23} which controls ν_μ and $\bar{\nu}_\mu$ disappearance, and to θ_{13} which controls ν_e appearance.

Both kinds of experiment are most sensitive to Δm^2_{31}, and so the (by definition) smaller difference Δm^2_{21} may be neglected in performing fits to these datasets.

We have not used data from *solar neutrino* observations, which measure the disappearance of low-energy electron neutrinos, and are dominantly sensitive to the "solar" mixing angle θ_{12}. Nor do we include data from *reactor neutrino* experiments, which are sensitive to both θ_{12} and θ_{13}.

3.2 Results from the two-flavor limit

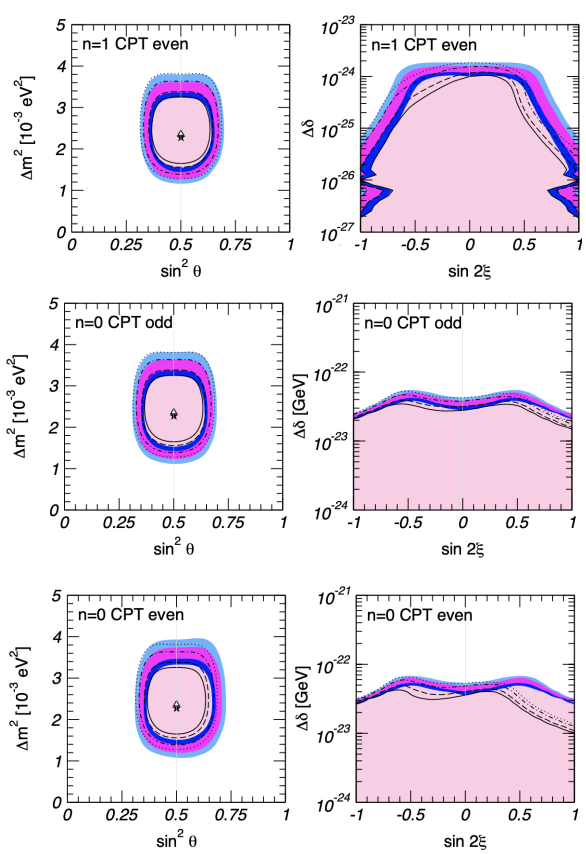
A natural place to start our work was with the easily-understood two-flavor oscillations described in section 1.3.6, where the parameters to fit are the parameters listed in table 1.1. So we began by attempting to reproduce the results of [1] with updated atmospheric neutrino data from Super-Kamiokande.

3.2.1 Reproducing prior results

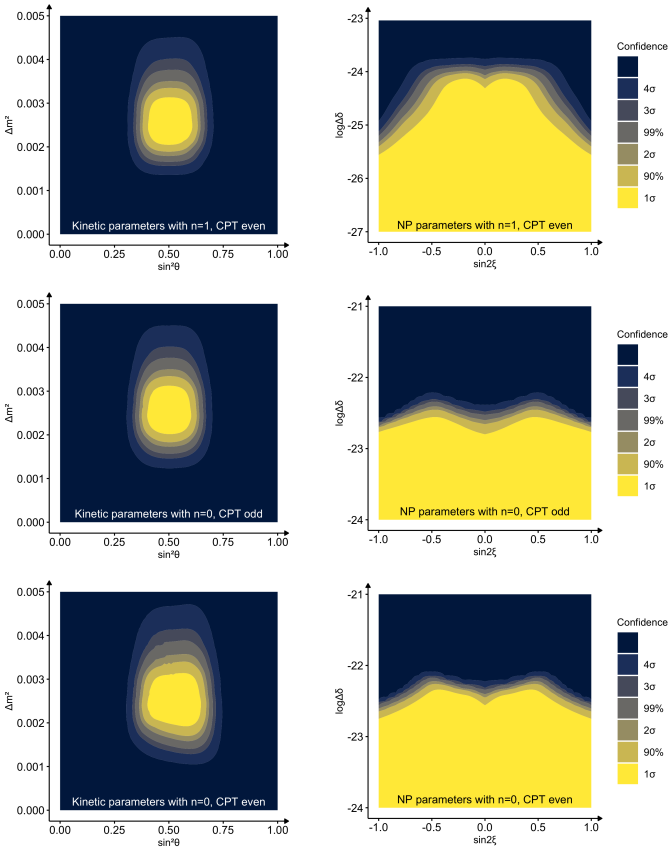
The previous work fitted these parameters to atmospheric neutrino data from Super-Kamiokande and long-baseline data from K2K. It wasn't possible to exactly reproduce this fit, since we didn't have the K2K data available to us in the two-flavor code. Instead, the plots in figure 3.1 below were obtained using only SK data. Partly compensating for the lack of K2K, we have *more* SK data than before.

In this and all following figures, Δm^2 is given in eV^2 and $\Delta\delta$ in GeV^{n+1}.

Figure 3.1: Comparison of excluded regions from two-flavor fits with SK data.



(a) Previous results borrowed directly from [1], which considered three cases, from top to bottom: n=1, CPT even; n=0, CPT odd; n=0, CPT even.



(b) Excluded regions from two-flavor fits to SK atmospheric data. Plots on left show confidence regions for kinetic parameters; plots on right show confidence regions for NP parameters.

The excluded regions we found are clearly very similar to those obtained by [1] in both shape and extent. Some differences were expected, given the differences in the datasets.

Of course, as a double-check, we fitted the data using three different optimization algorithms described in 2.2, and verified that the excluded regions were invariant under the choice of algorithm.

These results gave us confidence in the correctness of our numerical code and optimization algorithms.

3.2.2 Two-flavor oscillations with additional data

Even though we don't have the K2K data available, we do have other new data that was not available in 2004:

- we may include new atmospheric neutrinos data from DeepCore in our fit, and
- we may even include data from long-baseline accelerator experiments MINOS, NOνA, and T2K, which are sensitive to the same mass splitting and mixing angle as atmospheric neutrinos.

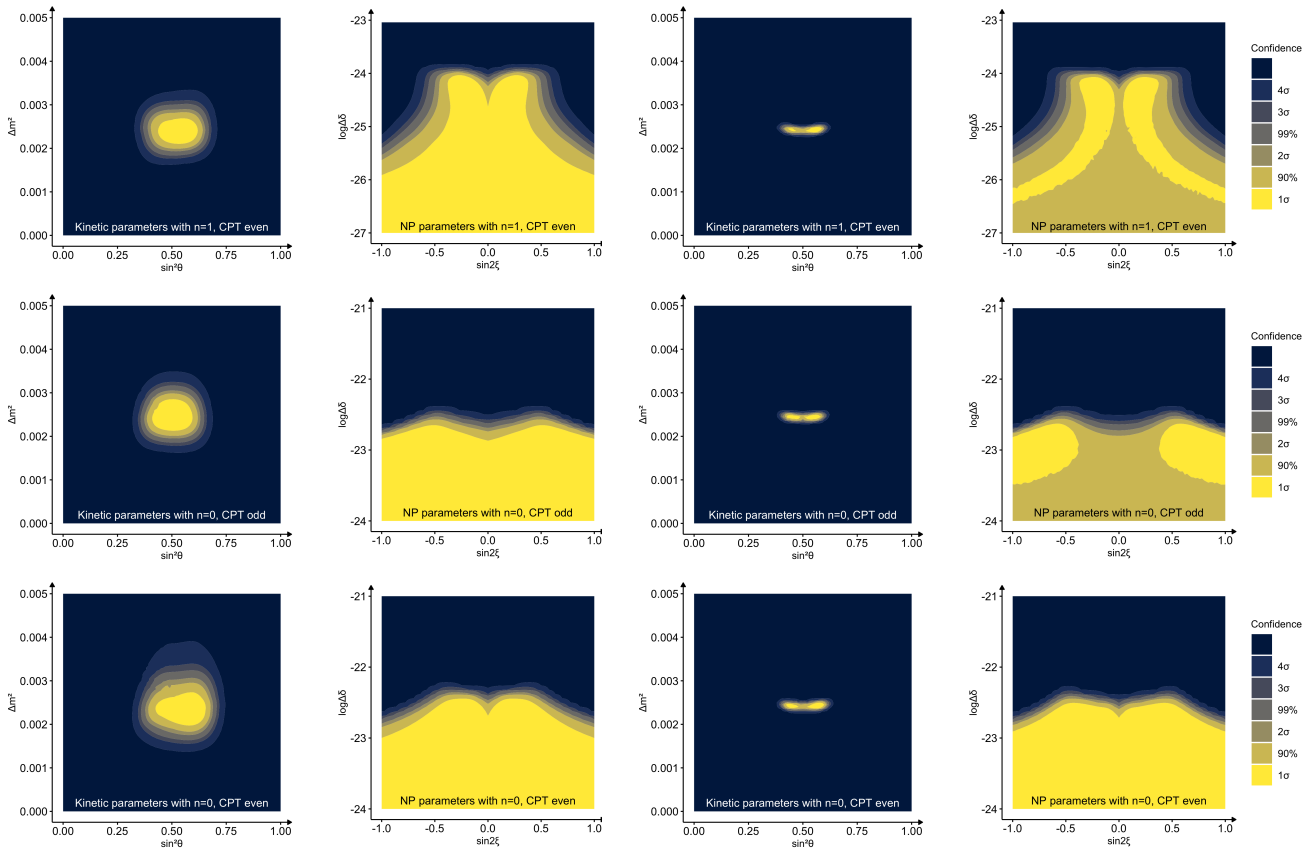
However, we encountered a subtlety here. When we initially included all the long-baseline data, we found that the raw χ^2 values of 2.1.2 produced were much larger than expected, indicating a poor model fit.

The explanation is that the T2K and NOνA experiments are sensitive not only to muon neutrino disappearance, but also to electron neutrino appearance, a phenomenon which cannot be explained in the two-flavor limit. But this doesn't affect nor undermine our estimates of the two-flavor NP parameters, since they can't explain electron neutrino appearance either. Therefore, the contribution of the appearance signal to χ^2 is just an overall constant offset, and we still obtain sensible and meaningful excluded regions from $\Delta\chi^2_{\text{profile}}$ where the constant offset cancels.

Figure 3.2 shows the results of such fits. It may be seen that the additional data provides much tighter constraints on the standard parameters θ and Δm^2, without much affecting the excluded region of the NP parameter space.

This result is expected for data from long-baseline accelerator experiments, where the energy of the neutrinos is lower than for atmospheric neutrinos. These experiments are therefore less sensitive to the NP effects we're looking for. Conversely, they have higher energy resolution and high statistics, which allows for a much more accurate determination of the standard kinetic oscillation parameters.

Figure 3.2: Two-flavor fits with SK and additional data from DeepCore, MINOS, NOνA, and T2K.



(a) Excluded regions from two-flavor fits with SK and DeepCore atmospheric data.

(b) Excluded regions from two-flavor fits with atmospheric data together with MINOS, NOνA, and T2K long-baseline accelerator data.

We used the same dataset to establish updated bounds on the magnitude of NP effects in the two-flavor limit.

3.2.3 Updated bounds on NP effects

To establish updated bounds on $\Delta\delta$, an additional, higher-resolution fit was performed on a one dimensional grid, with ξ treated as a nuisance parameter. The fit was performed twice, with two different datasets:

1. first, the fit was performed with data from all the experiments listed above: SL, DeepCore, and the long-baseline experiments, and then
2. the fit was re-run with the data from DeepCore excluded, to determine what impact it had on the results.

In table 3.1, the 3σ bounds are compared to the previous bounds reported in [1].

Overall, the bounds were improved by a factor of two. This modest improvement is not surprising, given that the SK data is already limited by systematic uncertainties, while, on the other hand, the DeepCore dataset is still limited in size. Nevertheless, we have verified that the inclusion of DeepCore data is relevant in the improvement of the bounds on the NP parameters. This is also understandable, since DeepCore is sensitive to higher-energy events. [1]

Table 3.1: Updated bounds on NP effects from two-flavor fits

NP scenario	Previous bound	New bound (DeepCore included)	New bound (DeepCore excluded)
$n = 1$, CPT even	1.6×10^{-24}	8.5×10^{-25}	1.0×10^{-24}
$n = 0$, CPT even	6.3×10^{-23}	2.7×10^{-23}	3.5×10^{-23}
$n = 0$, CPT odd	5.0×10^{-23}	2.4×10^{-23}	2.6×10^{-23}

3.3 Results from the hierarchical approximation

We're not yet ready to run fits with all the parameters describing three-flavor oscillations, since:

- these fits are extremely computationally expensive, and require execution in a clustered environment, and
- we don't yet have code to perform fits involving solar or reactor neutrinos, which are essential to constrain the Δm^2_{21} and the solar mixing angle θ_{12}, along with their NP friends $\Delta\delta_{21}$ and ξ_{12}.[2]

However, as a first step toward such computations, and as a validation of the code and methodology, we performed some computations with the parameters of the hierarchical approximation of section 1.3.8, that is, with just one additional parameter compared to the two-flavor limit above.

As mentioned above, some long-baseline experiments detect the appearance of electron neutrinos (and antineutrinos) in addition to measuring the disappearance of muon neutrinos. This appearance signal cannot be explained in the two-flavor limit used in the previous sections. But it can be explained when we add θ_{13} to the fit.

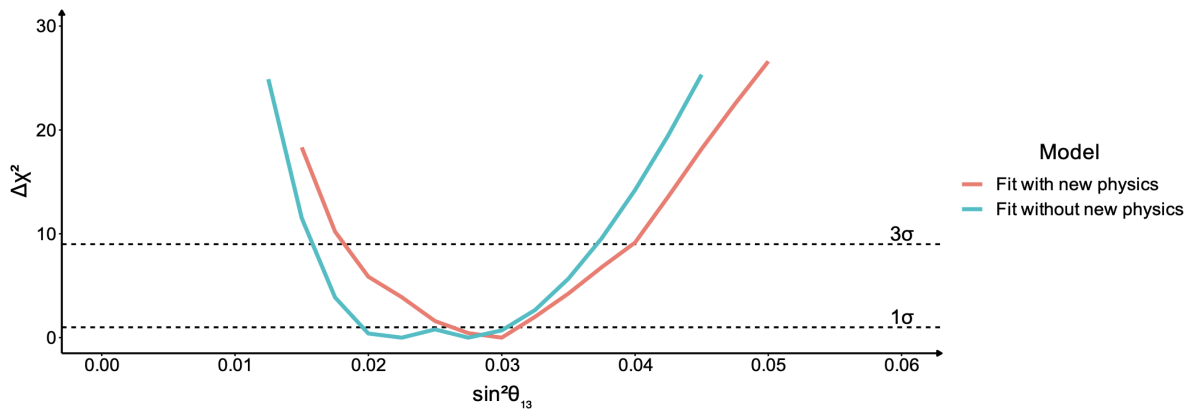
3.3.1 Estimating θ_{13} from long-baseline data

As a warmup, we estimate θ_{13} by fitting the long-baseline data with:

- only the kinetic parameters of the hierarchical approximation, but no new physics, and
- both the kinetic and NP parameters of the hierarchical approximation. Here we must minimize χ^2 over a 5-dimensional parameter space.

Figure 3.3 shows the dependence of $\Delta\chi^2$ on θ_{13}.

Figure 3.3: Dependence of $\Delta\chi^2$ on θ_{13} for fits with (red) and without (blue) new physics



[1] So we could expect a substantial improvement of the bounds with the upcoming data release from DeepCore.
[2] In fact, we don't yet even have code to perform three-flavor fits to the atmospheric neutrino data.

Intersections with the dashed horizontal lines determine the 1σ and 3σ confidence limits.

- The confidence interval for the scenario without new physics is consistent with the best present determination of this mixing angle from long-baseline experiments. For example, in [24], the T2K collaboration obtained $0.0268^{+0.0055}_{-0.0043}$ for normal ordering, or $0.0300^{+0.0059}_{-0.0050}$ for inverted ordering.
- Furthermore, the constraints on θ_{13} are only slightly relaxed in the scenario with new physics. This confirms that the data has sufficient information to meaningfully constrain the fitted parameters.

3.3.2 Constraints from ν_e appearance signal

We have argued above that the ν_e appearance signal is an important feature of the long-baseline experiments, that it is a feature which cannot be explained in the two-flavor limit, and that it *can* be explained by the introduction of θ_{13} to the fit. If all this is true, then it follows that excluding the appearance signal:

- should not affect the confidence regions obtained from a two-flavor parameterization, but
- should cause problems for a fit to the parameters of the hierarchical approximation, where one of the degrees of freedom would be insufficiently constrained.

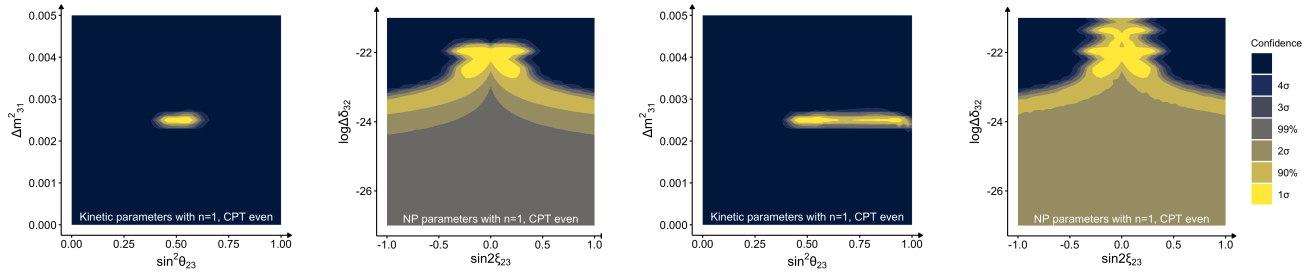
This allows for a useful test of our model and code.

The following figures show the results of fitting:

- with and without θ_{13} as a free parameter to be estimated, and
- with and without the appearance signal.

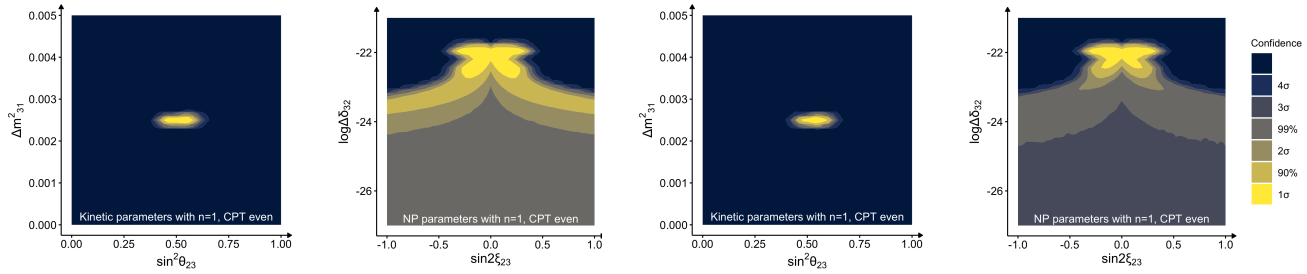
For the CPT even case with $n = 1$ we obtain the fits shown in figure 3.4, where the difference between the lower and upper plots arises from the inclusion or exclusion of the ν_e appearance signal, and only the four plots on the right include θ_{13} as a free parameter to be estimated.

Figure 3.4: Fits with CPT even, $n = 1$ to disappearance and appearance signals in MINOS, NOνA, and T2K long-baseline accelerator data.



(a) Excluded regions from fit with $\theta_{13} = 0$ to disappearance signal.

(b) Excluded regions from fit with θ_{13} free to disappearance signal.



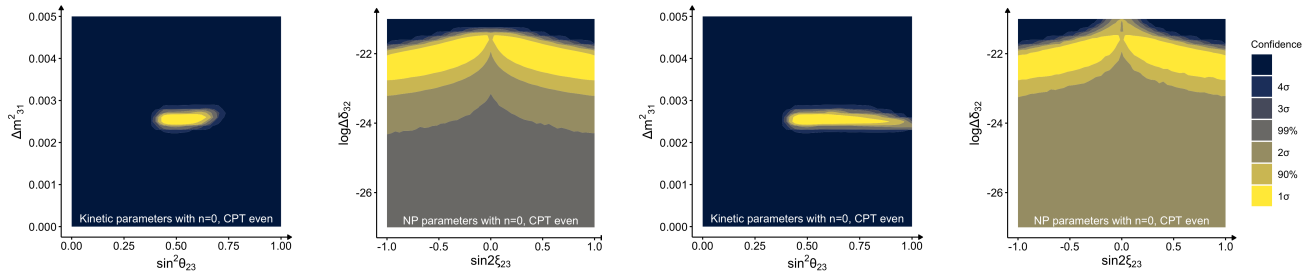
(c) Excluded regions from fit with $\theta_{13} = 0$ to both disappearance and appearance signals taken together.

(d) Excluded regions from fit with θ_{13} free to both disappearance and appearance signals taken together.

These plots confirm our predictions, showing that exclusion of the appearance signal does not affect the confidence regions obtained under the two-flavor parameterization, but that the appearance signal plays a very important role in constraining both the kinetic parameters and the NP parameters as soon as we introduce θ_{13} into the fit.

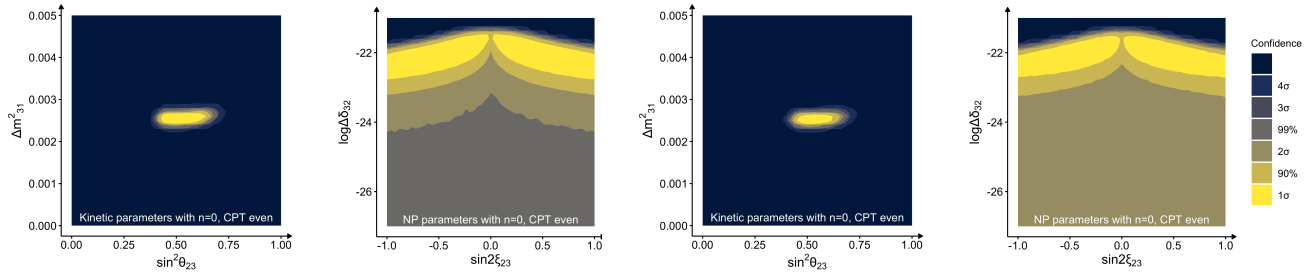
For the CPT even and odd cases with $n = 0$ we obtain the fits shown in figures 3.5 and 3.6.

Figure 3.5: Fits with CPT even, $n = 0$ to disappearance and appearance signals in MINOS, NOνA, and T2K long-baseline accelerator data.



(a) Excluded regions from fit with $\theta_{13} = 0$ to disappearance signal.

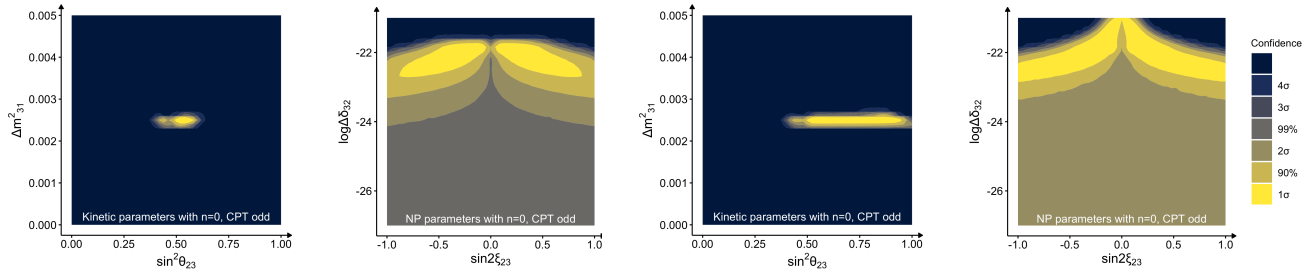
(b) Excluded regions from fit with θ_{13} free to disappearance signal.



(c) Excluded regions from fit with $\theta_{13} = 0$ to both disappearance and appearance signals taken together.

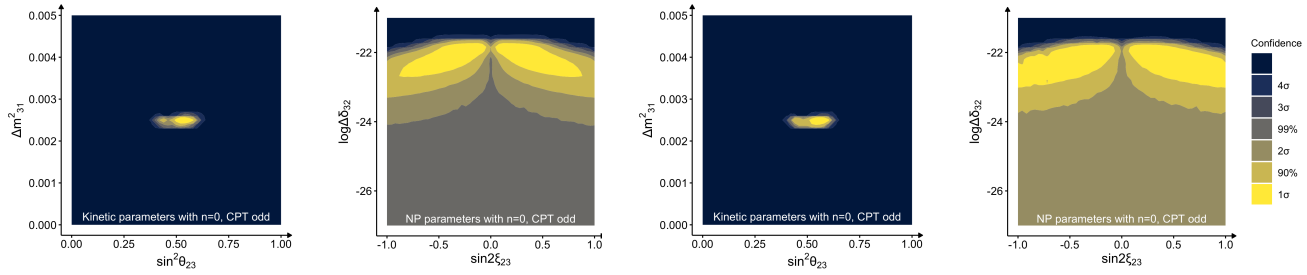
(d) Excluded regions from fit with θ_{13} free to both disappearance and appearance signals taken together.

Figure 3.6: Fits with CPT odd, $n = 0$ to disappearance and appearance signals in MINOS, NOνA, and T2K long-baseline accelerator data.



(a) Excluded regions from fit with $\theta_{13} = 0$ to disappearance signal.

(b) Excluded regions from fit with θ_{13} free to disappearance signal.



(c) Excluded regions from fit with $\theta_{13} = 0$ to both disappearance and appearance signals taken together.

(d) Excluded regions from fit with θ_{13} free to both disappearance and appearance signals taken together.

In both cases the confidence regions are consistent with our predictions.

We now draw attention to a particular feature of the confidence interval for $\sin^2\theta_{23}$ when θ_{13} is included in the fit to the ν_μ disappearance data, with the ν_e appearance signal excluded. In this under-constrained fit, the range for $\sin^2\theta_{23}$ extends to its maximum allowed value.

This can be understood by considering this approximate analytic expression for the ν_μ survival probability:

$$P_L(\mu \to \mu) \simeq 1 - \sin^2 2\theta_{\mu\mu} \sin^2\left(\frac{1}{4}\frac{L}{E}\Delta m^2_{23}\right) \qquad \sin^2\theta_{\mu\mu} \equiv \cos^2\theta_{13}\sin^2\theta_{23} \tag{3.1}$$

Setting $\theta_{13} = 0$ as in the two-flavor limit leads to $\theta_{\mu\mu} = \theta_{23}$, and we recover equation (1.40).

Now, given a constraint from data on the survival probability, if θ_{13} is nonzero then $\cos^2\theta_{13} < 1$, and so $\sin^2\theta_{23}$ must increase to compensate. This is precisely what we observe in subfigures 3.4b, 3.5b, and 3.6b.

On the other hand, with the ν_e appearance signal *included*, θ_{13} is constrained to the small value found in 3.3.1 above, such that $\cos^2\theta_{13}$ is very close to unity, and so the confidence interval for $\sin^2\theta_{23}$ stays close its usual estimate in subfigures 3.4d, 3.5d, and 3.6d.

The plots also show that the inclusion of θ_{13} does not much affect our bounds on the NP parameters when the appearance signal is present, suggesting that our bounds obtained above in 3.2.3 are at least somewhat robust. On the other hand, we are unsurprised to observe that exclusion of the atmospheric data results in overall looser bounds on the strength of the NP effect.

3.4 Summary

Using a parameterization of the two-flavor limit to model muon neutrino disappearance, we have reproduced the previous work reported in [1] and updated the bounds reported there, obtaining a modest (factor of two) improvement in the constraint on $\Delta\delta_{31}$ via the incorporation of new data from SK, DeepCore, MINOS, NOνA, and T2K.

We have taken a first step toward generalizing these results to a full parameterization of three-flavor neutrino oscillations, by establishing the infrastructure for performing the much more computationally intensive fits required. Our evaluation of a range of standard optimization algorithms concluded that the simplex method exhibited slightly better performance than population methods for our particular objective function, at least for optimization over a three or four dimensional parameter space.

The infrastructure for fitting such three-flavor parameterizations was validated under the hierarchical approximation, which introduced θ_{13} as a free parameter. We showed that our algorithms were able to obtain the accepted value of θ_{13} from the long-baseline accelerator data, and we saw the importance of the electron neutrino appearance signal in constraining the fitted parameters. On the other hand, we also confirmed that the inability of the two-flavor limit to account for the appearance signal did not render invalid the constraints on NP parameters which were obtained by assuming this limit to hold.

In the next phase of this work we will introduce data from solar neutrinos and reactor experiments and fit a full parameterization of three-flavor mixing. This is a far more computationally-intensive task due to the extra dimensionality of the parameter space.

Bibliography

[1] MC Gonzalez Garcia and Michele Maltoni. "Atmospheric Neutrino Oscillations and New Physics". In: *Phys.Rev.D* 70.033010 (2004). DOI: https://doi.org/10.48550/arXiv.hep-ph/0404085.

[2] MC Gonzalez Garcia and Michele Maltoni. "Phenomenology with Massive Neutrinos". In: *Phys.Rept* 460.1-129 (2008). DOI: https://doi.org/10.48550/arXiv.0704.1800.

[3] DS Harmer RJ Davis and K Hoffman. "Search for Neutrinos from the Sun". In: *Phys.Rev.Lett* 20 (1968). DOI: https://doi.org/10.1103/PhysRevLett.20.1205.

[4] Carlo Giunti and Chung W Kim. *Neutrino Physics and Astrophysics*. Oxford University Press, 2007.

[5] L Wolfenstein. "Neutrino oscillations in matter". In: *Phys.Rev.D* 17 (1978). DOI: https://doi.org/10.1103/PhysRevD.17.2369.

[6] V De Sabbata and M Gasperini. "Neutrino oscillations in the presence of torsion". In: *Il Nuovo Cimento A* 65 (1981). DOI: https://doi.org/10.1007/BF02902051.

[7] Sidney Coleman and Sheldon L Glashow. "Cosmic Ray and Neutrino Tests of Special Relativity". In: *Phys.Lett.B* 405 (1997). DOI: https://doi.org/10.48550/arXiv.hep-ph/9703240.

[8] M Gasperini. "Testing the principle of equivalence with neutrino oscillations". In: *Phys.Rev.D* 38 (1988). DOI: https://doi.org/10.1103/PhysRevD.38.2635.

[9] Don Colladay and V Alan Kostelecký. "CPT violation and the standard model". In: *Phys.Rev.D* 55 (1997). DOI: https://doi.org/10.1103/PhysRevD.55.6760.

[10] S Pakvasa and JWF Valle. "Neutrino Properties Before and After KamLAND". In: *Proc.Indian Natl.Sci.Acad* 70A (2004). DOI: https://doi.org/10.48550/arXiv.hep-ph/0301061.

[11] S Lavignac C Giganti and M Zito. "Neutrino oscillations: The rise of the PMNS paradigm". In: *Progress in Particle and Nuclear Physics* 98 (2018). DOI: 10.1016/j.ppnp.2017.10.001.

[12] M. C. Gonzalez-Garcia, Michele Maltoni, and Jordi Salvado. "Testing matter effects in propagation of atmospheric and long-baseline neutrinos". In: *Journal of High Energy Physics* 2011.5 (2011). DOI: https://doi.org/10.48550/arXiv.1103.4365.

[13] RL Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[14] Steve Baker and Robert D Cousins. "Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms". In: *Nucl.Instrum.Meth.* 221 (1984). DOI: https://doi.org/10.1016/0167-5087(84)90016-4.

[15] Mykel J Kochenderfer and Tim A Wheeler. *Algorithms for optimization*. The MIT Press, 2019.

[16] Y Nesterov. "A Method of Solving a Convex Programming Problem with Convergence Rate $O(1/k_2)$". In: *Soviet Mathematics Doklady* 27 (1983).

[17] D Kingma and J Ba. "Adam: A Method for Stochastic Optimization". In: *International Conference on Learning Representations* (2015). DOI: https://doi.org/10.48550/arXiv.1412.6980.

[18] JA Nelder and R Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7 (1965). DOI: https://doi.org/10.1093/comjnl/8.1.27.

[19] et al Jeffrey C Lagarias. "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions". In: *SIAM J.OPTIM.* 9 (1998). DOI: https://doi.org/10.1137/S1052623496303470.

[20] Rainer Storn and Kenneth Price. "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces". In: *Journal of Global Optimization* 11 (1997). DOI: https://doi.org/10.1023/A:1008202821328.

[21] S Das and PN Suganthan. "Differential Evolution: A Survey of the State-of-the-Art". In: *EEE Transactions on Evolutionary Computation* 15 (2011). DOI: https://doi.org/10.1109/TEVC.2010.2059031.
[22] RC Eberhart J Kennedy and Y Shi. *Swarm Intelligence*. Morgan Kaufmann, 2001.
[23] Gregory D Martinez et al. "Comparison of statistical sampling methods with ScannerBit, the GAMBIT scanning module". In: *Eur.Phys.J.C* (2017). DOI: https://doi.org/10.48550/arXiv.1705.07959.
[24] T2K Collaboration. "Improved constraints on neutrino mixing from the T2K experiment with 3.13×10^{21} protons on target". In: *Phys.Rev.D* 103.11 (2021). DOI: 10.1103/physrevd.103.112008.