Computational Physics Project 1, Autumn 2019

S. S. Mall

***Abstract*—The survival probability of a muon neutrino is governed by its energy, and the mixing parameters θ_{23} and Δm^2_{23}. By considering independent measurements of the number of neutrino events observed for binned energies compared to the expected rate given by the Poisson distribution, a Negative Log Likelihood (NLL) function was found. The minimisation of this function leads to the best estimate for the mixing parameters.**

Univariate, gradient descent, and quasi-Newton methods were used to find the global minimum of the NLL, leading to a value of $\theta_{23} = 0.8589 \pm 0.01555$ and $\Delta m^2_{23} = (2.885 \pm 0.02862) \times 10^{-3}$ (eV)^2. The NLL at this point is 787.2.

Finally, the cross-section of the interaction was considered, leading to the survival probability increasing with energy by a constant of proportionality, *c*. The quasi-Newton method was used to minimise with respect to all three parameters, leading to a final position of $\theta_{23} = 0.8201 \pm 0.0223$, $\Delta m^2_{23} = (3.038 \pm 0.0265) \times 10^{-3}$ (eV)^2, and $c = 1.636 \pm 0.0612$ (GeV)$^{-1}$. The NLL at this point is 341.2.

I. Introduction and Background

Neutrinos are fundamental particles with three lepton 'flavours' denoted by an integer from one to three describing their generation (electron, muon, and tauon respectively). Although the Standard Model predicts that the neutrino flavour states and energy eigenstates are equal, it is found that the flavour states can be expressed as a linear sum of energy states governed by a mixing angle. This leads to a time-dependence in the probability that a neutrino will collapse into a specified flavour state, i.e. it may not collapse in to the state that it was prepared in [1].

The probability of a $\nu_\mu \rightarrow \nu_\mu$ oscillation, $P_{\nu_\mu \rightarrow \nu_\mu}(E, \theta_{23}, \Delta m^2_{23})$ is given in Eqn 1 as a function of the neutrino energy, and the mixing parameters associated with oscillations between the muon flavour state and the electron flavour state: θ_{23} is the mixing angle and $\Delta m^2_{23} = m^2_3 - m^2_2$ is the mass-square-difference. This is a first order approximation that neglects the $1,3$ oscillation by assuming θ_{13} is small [2].

The given data is the number of observed muon neutrinos for each of 200 bins of energy from 0 to 10 GeV. The number of events in each bin follows a Poisson distribution with a mean that was also given (this time found by simulation). However, this mean value does not consider the probability of the neutrino oscillating to different flavours. To consider this, the given rate has to be multiplied by the probability of $\nu_\mu \rightarrow \nu_\mu$ oscillation, $P_{\nu_\mu \rightarrow \nu_\mu}$ [1]:

$$P_{\nu_\mu \rightarrow \nu_\mu} = 1 - \sin^2(2\theta_{23}) \sin^2\left(\frac{1.267 \Delta m^2_{23} L}{E}\right) \quad (1)$$

Assuming the measurements are independent, each energy bin i follows the distribution

$$\mathcal{P}_i(k_i\ ;\ \theta_{23}, \Delta m^2_{23}) = \frac{\lambda_i(\theta_{23}, \Delta m^2_{23})^{k_i}\ .\ \exp^{-\lambda_i(\theta_{23}, \Delta m^2_{23})}}{k_i!} \quad (2)$$

where

$$\lambda_i(\theta_{23}, \Delta m^2_{23}) = \tilde{\lambda_i}\ .\ P_{\nu_\mu \rightarrow \nu_\mu}(E_i, \theta_{23}, \Delta m^2_{23}) \quad (3)$$

and $\tilde{\lambda}_i$ is the simulated rate for energy bin i. The energy of bin i is the centre of the bin.

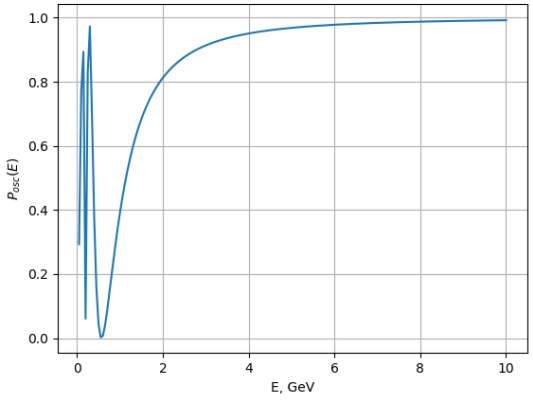

Fig. 1. The probability of muon-to-muon neutrino oscillation as a function of energy. The mixing parameters are held constant at $\theta_{23} = \frac{\pi}{4}$ and $\Delta m^2_{23} = 2.4 \times 10^{-3}$ (eV)^2.

The fit parameters θ_{23} and Δm^2_{23} can be found by maximising the Likelihood or conversely, by minimising the Negative Log Likelihood (NLL). Because the energy bins are treated as independent measurements, the likelihood is the product of the Poisson distribution of all energy bins. Estimates for the fit parameters are the parameters for which the NLL (Eqn. 4) is minimised.

$$\text{NLL}(\theta_{23}, \Delta m^2_{23}) = -\sum_i \log\left(\mathcal{P}_i(k_i\ ;\ \theta_{23}, \Delta m^2_{23})\right) \quad (4)$$

II. Method Overviews

A. Parabolic Method

The parabolic method minimises a function along one direction. The function is passed 3 guess points (and in this case, the value of the variable held constant, e.g. the mass-square-difference if minimising along the mixing angle) which it uses to approximate the NLL function as a parabola. This is justified if the points are sufficiently close to the function's

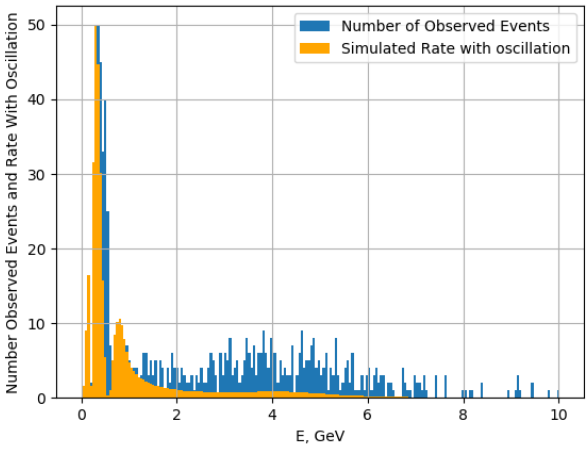

Fig. 2. The mean predicted rate with neutrino oscillation considered (orange), compared to the observed rate (blue). Each energy bin is distributed according to the Poisson distribution. Because the rate is a function of the mixing parameters, the number of observed events can be used to extract its fit parameters. Here the oscillation probability is evaluated at $\theta_{23} = \frac{\pi}{4}$, $\Delta m^2_{23} = 2.4 \times 10^{-3}$ (eV)^2.

minimum and becomes more accurate closer to the minimum point.

The approximation is made by fitting a second order Lagrange Polynomial to the three points and the minimum point of the parabola is easily found as a function of the three input points. Of the four points, the point with the largest NLL is discarded and the remaining points sorted and used to fit the next parabola. This is iterated until a convergence condition is satisfied and the minimum point of the most recently fitted parabola is returned as the minimum point of the function.

The convergence condition here is simply that the change in the minimum point between the current and previous parabola is smaller than 10^{-8}. Although this is very crude, it is justified as this threshold value is several orders of magnitude lower than the uncertainty in the minimum point (as shown in section III).

B. Univariate Minimiser

The univariate minimiser is a two dimensional minimiser that carries out a one dimensional minimisation along each direction in series, repeating until a convergence condition is met. The one-dimensional minimiser used here is the parabolic method.

The univariate function is passed three points per variable to fit an initial parabola with the second point assumed to be closest to the minimum (the variables not being minimised will be held constant at this value during the first minimisation). While the convergence condition is not met, a parabolic minimisation is carried out along each direction successively with the other variables held constant at their previous minimum point.

It should be noted here that the parabolic minimiser takes three input parameters (to fit the initial parabola) but only returns one value (the minimum point) so the parabolic minimiser was modified to also return the three points used to fit the final parabola. These were used as the input points for the next minimisation along that direction.

C. Gradient Descent Method

The gradient descent method takes a single value for each variable as an initial guess and calculates the gradient vector at this point. This vector is the direction of greatest ascent so by considering points in the opposite direction, the minimiser will move closer to the minimum point with each iteration.

This requires each step to be small enough so that the gradient does not change significantly so a parameter $\underline{\alpha}$ is introduced to ensure this. Each successive step in the direction i is given by:

$$x_{n+1,i} = x_{n,i} - \alpha_i \, \nabla_i f(\underline{x_n}) \tag{5}$$

$\underline{\alpha}$ is a vector as the function changes on different scales with respect to each parameter – this parameter is passed to the Gradient Descent function by the user and requires some tuning or knowledge of the function beforehand.

The gradient was calculated numerically using the central finite difference scheme.

D. Quasi-Newton Gradient Method

Newton's Method follows a similar approach to gradient descent, but also considers the curvature of the function at the point $\underline{x_n}$ by acting the inverse Hessian matrix $\mathbf{H}_n^-1$ on the vector $\alpha_i \, \nabla_i f(\underline{x_n})$ in Eqn 5.

Because matrix inversion can be computationally expensive for higher dimensions, the quasi-Newton method uses the matrix $\mathbf{G}_n$ which satisfies

$$\boldsymbol{G}_n \,.\, \underline{a} \approx \boldsymbol{H}_n^{-1} \,.\, \underline{a} \tag{6}$$

for some vector $\underline{a}$.

A recurrence relation between $\mathbf{G}_{n+1}$ and $\mathbf{G}_n$ is given by the DFP algorithm [3], where $\mathbf{G}_0$ is the identity matrix and $\mathbf{G}_{n+1}$ is also a function of the n^{th} and $(n+1)^{\text{th}}$ points.

Given that x_n and $\mathbf{G}_n$ are known, the order of calculations is to first find $\underline{\nabla} f(\underline{x_n})$ and use these to find the $(n+1)^{\text{th}}$ position and gradient. Now that the position and gradient of both the n^{th} and $(n+1)^{\text{th}}$ points are known, $\mathbf{G}_{n+1}$ can be found and the process is iterated until a convergence condition is satisfied.

The quasi-Newton function was written to accept any function of two or three variables.

E. Convergence Condition

The convergence condition for all two and three dimensional algorithms was for the ratio between the function at the current and previous position must be sufficiently close to 1, i.e.

$$0 \; < \; 1 - \frac{f(\underline{x}_n)}{f(\underline{x}_{n-1})} \; < \; 10^{-8} \tag{7}$$

This ensures that the most recent change in the function is small, and also that the final value of the function is lower than the previous.

III. Uncertainty in Results

A. Curvature Method

The second order Taylor expansion of an NLL function of one variable around the minimum $\hat{x}$ is given by

$$\log L(x) \approx \log L(\hat{x}) - \frac{1}{2}\frac{d^2}{dx^2}\big(\log L(\hat{x})\big) \,.\, (x-\hat{x})^2 \tag{8}$$

.

Defining

$$\frac{1}{\sigma^2} = \frac{d^2}{dx^2}\big(\log L(\hat{x})\big) \tag{9}$$

it can be seen that the Likelihood function around the minimum point is a Gaussian of width σ_x. This is the uncertainty in $\hat{x}$ and can be generalised to higher dimensions as the inverse Hessian matrix.

The second derivative with respect to a single variable was found using a central finite difference scheme with an ϵ value of 10^{-7}.

B. Parabolic Approximation

The curvature of the approximating parabola can be found using the functional form of the second order Lagrange polynomial $P_2(x)$ and collecting the factors in front of the x^2 terms. The uncertainty is then found using Eqn 9.

C. Gaussian Approximation

The NLL can be also approximated to a Gaussian close to the minimum point, in which case the NLL evaluated at a distance of $\pm\sigma_x$ from the minimum point is the minimum NLL plus an absolute value of $\frac{1}{2}$.

In each iteration a point is considered above and below the minimum, increasing in distance from it by a step size determined by the user (as opposed to iterating above then below in series). Values for σ^+ and σ^- are returned when both points are within a user-defined difference to $\text{NLL}_{\text{min}} + 0.5$.

IV. Testing

A. 1d Parabolic Minimisation

The NLL function was minimised along θ_{23} while holding Δm^2_{23} constant at a value of 2.4×10^{-3} (eV)^2. Using Fig 3, the initial guess points were $(0.8, 0.9, 0.95)$ and the function converged within 8 steps to $\theta_{23} = 0.9530$ with the NLL at this point equal to 871.3.

The uncertainty in this value was calculated using all three methods and displayed in Table I and the agreement of the values used as an indicator of the accuracy of the approximations made.

Because the error using the Gaussian approximation is asymmetric, the total uncertainty is quoted as $\frac{\sigma^+ - \sigma^-}{2}$, where $\sigma_\pm$ was found to be (0.9429 , 0.9633) and the step size and threshold were both 10^{-5}. It can be seen from Fig. 3 that this is a reasonable uncertainty.

It can be seen that the values agree with each other to the second significant figure and only differ by 1 in the third. This suggests that the parabolic approximation made by the parabolic minimiser is valid as it agrees with the curvature method which makes no such approximation.

Curvature	Parabolic	Gaussian
0.010185	0.010169	0.010125

TABLE I
A table of the uncertainty values in the minimum value of θ_{23}

It is expected that as θ_{23} approaches $\frac{\pi}{4}$, the Gaussian and parabolic approximations become less accurate and hence the uncertainty values will diverge from each other.

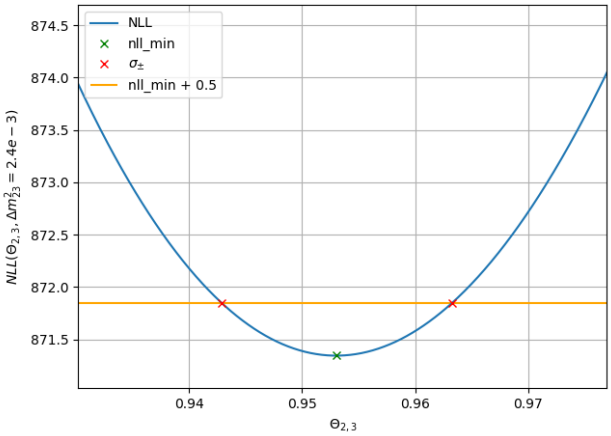

Fig. 3. A one-dimensional minimisation of NLL($\theta_{23}, \Delta m^2_{23}$). The value of Δm^2_{23} is held constant at 2.3×10^{-3} eV^2. The asymmetric uncertainty is found using the Gaussian approximation where the NLL shifts from its minimum value by 0.5. It can be seen that the algorithm produced $\sigma_\pm$ reasonably close to this point.

B. Univariate

The two-dimensional minimisation was carried out using the univariate, gradient descent, and quasi-Newton methods, while the uncertainty was found using only the curvature method (assuming the off-diagonal elements of the Hessian are negligible). Firstly however, tests were run to determine how each minimiser behaved under different conditions.

From Fig. 4, it can be expected that minimising along Δm^2_{23} first will converge faster to the global minimum than along θ_{23} because in the second case, the univariate will first minimise to $\theta_{23} \approx 0.8$ and then to one of the other local minima. The initial points must be chosen so that convergence is to the global minimum.

Furthermore, it is found that minimising with respect to Δm^2_{23} first required fewer iterations than the second case (16 and 22 iterations respectively), though this may not be the general case due to the relatively low number of steps.

Regarding sensitivity to the initial position, the univariate was initialised at four different initial positions and the results shown in Fig. 5. All four tests converged to the expected positions: two minima roughly symmetrical about $\theta_{23} = \frac{\pi}{4}$ with $\Delta m^2_{23} \approx 2.9 \times 10^{-3}$.

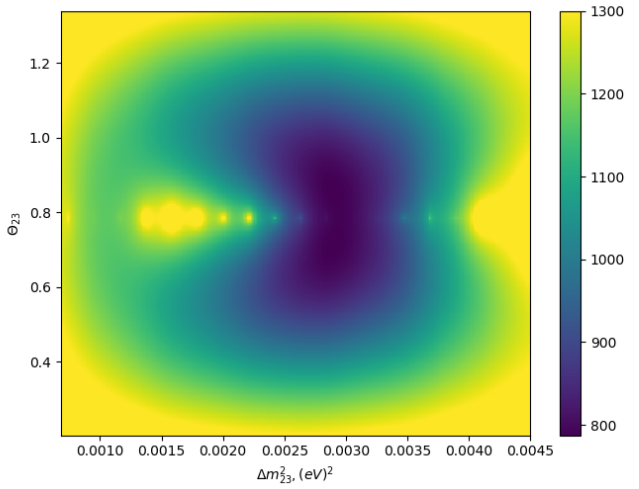

Fig. 4. The NLL as a function of mixing angle and mass-square-difference over a small range. This can be minimised for an estimate of the parameters.

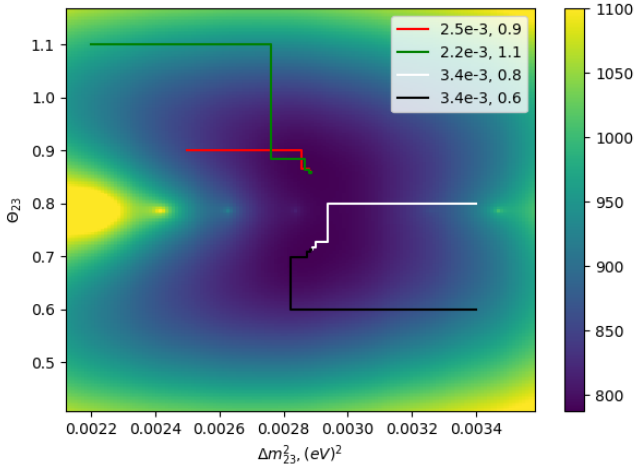

Fig. 5. The univariate was tested with four different initial positions $(\Delta m^2_{23}, \theta_{23}$, given by the values in the legend. There is one minimum point above and one below $\theta_{23} = \frac{\pi}{4}$ so the minimiser is expected to converge to one of these points depending on if it is initialised above or below this line. This is clearly seen in the figure.

1) Gradient Descent and Quasi-Newton: The gradient descent and quasi-Newton methods were validated using basic functions (a function of two variables for both methods and a function of three variables for the quasi-Newton method) with known minimum points. The functions used were:

$$f(x, y) = x^2 + y^2 \tag{10}$$

and

$$f(x, y, z) = x^2 + y^2 + z^2 + xy \tag{11}$$

These functions have minimum points at the origin and the characteristics of the **G** matrix can be easily deduced as it is related to the function's Hessian matrix.

Firstly, both minimisers tend towards the origin which is as expected, however due to the convergence condition given by Eqn 7 and that the function at the origin is equal to zero, the minimisers are not able to return a numerical result. However the test is considered successful as this is accounted for and is unlikely to occur for the NLL function.

Secondly considering the test function 11, the second derivatives with respect to each variable shows the the Hessian matrix must be symmetric with a constant diagonal of the same value, and the only non-zero off-diagonal terms are in the position of the $\frac{\partial^2}{\partial x \partial y}$, (the position 1,2) which is also constant. The inverse of this matrix has zero and non-zero terms in the same positions.

An example of the $\mathbf{G}_n$ matrix for an iteration while testing the function is given below. While it does not agree entirely with the prediction because the diagonal elements are not equal, the zero and non-zero elements are in the predicted positions.

$$\begin{bmatrix} 2.185 & 0.248 & 0. \\ 0.248 & 1.689 & 0. \\ 0. & 0. & 1. \end{bmatrix}$$

V. Results and Discussion

A. Two Dimensional Minimisation

Once the univariate, gradient descent, and quasi-Newton methods were validated, they were all initialised at the position $(\Delta m^2_{23}, \theta_{23} = 2.2 \times 10^{-3}\ ,\ 1.10)$ and the numerical results compared in Table II. The α parameter for the gradient descent and quasi-Newton methods was $(10^{-9}, 10^{-4})$.

	Univariate	Gradient Descent	Quasi-Newton
θ_{23}	0.8595	0.8585	0.8586
σ_θ	0.01557	0.01554	0.01553
$\Delta m^2_{23} \times 10^{-3}$	2.885	2.885	2.885
$\sigma_m \times 10^{-3}$	0.02861	0.02862	0.02862
No. Iterations	16	27	22

TABLE II
A table of results of a 2D minimisation using the univariate, gradient descent, and quasi-Newton methods.

The table suggests that all three methods agree very well. This is also backed up by Fig. 6, which shows the path taken by each method. All three terminate at the points given in Table V-A and the gradient descent and quasi-Newton follow a very similar path. The NLL at this point is 787.2.

In this case, a second check can be made. The quasi-Newton method is equivalent to the gradient descent if all $\mathbf{G}_n$ are unitary. Because the paths taken by both methods are very similar, it can be predicted that the $\mathbf{G}_n$ is close to unitary. This is found to be the case, as the final $\mathbf{G}_n$ matrix is found to be:

$$\begin{bmatrix} 1.005 & 7.053 \times 10^{-6} \\ 7.053 \times 10^{-6} & 1.000 \end{bmatrix}$$

This may be because the NLL function is close to parabolic over this range so the quasi-Newton method does not provide a distinct advantage in this case.

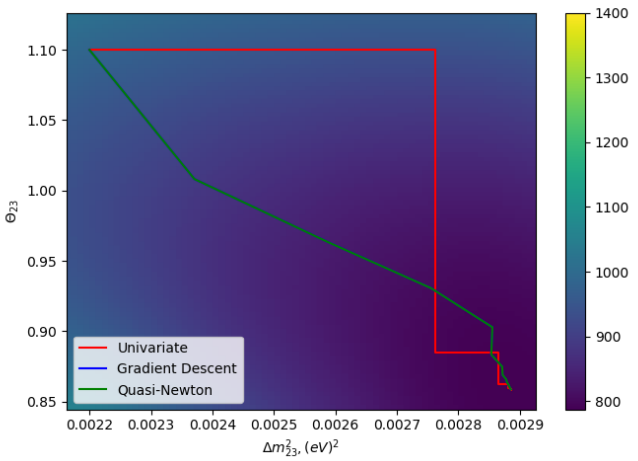

Fig. 6. The paths taken by three different minimisers from the same initial point. All three converge to roughly the same point (within one sigma uncertainty in each direction), with the gradient descent and quasi-Newton methods taking an almost identical path.

B. Three Dimensional Minimisation

The cross-section of the neutrino interaction means the rate of oscillation increases linearly with energy. The formula for oscillation probability is multiplied by a linear function of energy, i.e. Eqn 3 is modified so that

$$\lambda_i(\theta_{23}, \Delta m^2_{23}, c) = \tilde{\lambda_i} \, . \, P_{\nu_\mu \to \nu_\mu}(E_i, \theta_{23}, \Delta m^2_{23}) \, . \, c \, E_i \quad (12)$$

where c is a constant of proportionality. A modified NLL function is written to consider this parameter and the quasi-Newton method is used to minimise the new function with respect to all three variables.

Several starting points were tested and the point that converged to the lowest NLL was ($\theta_{23} = 0.8$, $\Delta m^2_{23} = 2.8 \times 10^{-3}$, $c = 1.6$) with $\underline{\alpha} = (10^{-4}, 10^{-9}, 10^{-3})$. The method converged to a minimum point given by Table III with the uncertainties found using the curvature method. The NLL at this point is 341.2, which is significantly lower than the two dimensional result. This suggests that the rate of neutrino oscillation does indeed increase linearly with energy.

VI. Conclusion

A NNL function was constructed from the Poisson distribution of each energy bin with a mean rate as a function of the mixing parameters and bin energy.

The minimum of the NLL function with respect to the mixing parameters was found using the univariate, gradient descent, and quasi-Newton methods. All three methods converged to the same minimum point and had very similar uncertainties. The final parameters and uncertainties (averaged over the three methods) are $\theta_{23} = 0.8589 \pm 0.01555$, $\Delta m^2_{23} = (2.885 \pm 0.02862) \times 10^{-3}$ (eV)^2. The NLL at this point is 787.2.

Secondly, the rate was assumed to increase with energy by a constant of proportionality c related to the interaction cross-section. The quasi-Newton method was used to minimise the new NLL function with respect to all three parameters, giving a new minimum point of $\theta_{23} = 0.8201 \pm 0.0223$, $\Delta m^2_{23} = (3.038 \pm 0.0265) \times 10^{-3}$ (eV)^2, and $c = 1.636 \pm 0.0612$ (GeV)$^{-1}$. The NLL at this point is 341.2. The lower NLL suggests that Eqn 12 is closer to the true rate than Eqn 3.

VII. Bibliography

1) H. Araujo (2018). Imperial College London, Department of Physics. Nuclear and Particle Physics (Yr 2) Course Notes, page 66.
2) RAUT, S. (2013). EFFECT OF NONZERO θ_{13} ON THE MEASUREMENT OF θ23. Modern Physics Letters A, 28(20), p.1350093.
3) Póczos, B. and Tibshirani, R. (n.d.). Convex Optimization CMU-10725 Quasi Newton Methods. [online] Stat.cmu.edu. Available at: https://www.stat.cmu.edu/~ryantibs/convexopt-F13/lectures/11-QuasiNewton.pdf [Accessed 16 Dec. 2019].