# Extracting Neutrino Oscillation Parameters from a Log-Likelihood Fit - Project 1

Shaun Fendi Gan, (Word Count: 2481)

***Abstract*—Negative Log-Likelihood (NLL) for the survival probability of neutrino oscillations was minimised to find model values that described the physical phenomenon. A 2D Univariate parabolic minimiser and N-dimensional Simulated Annealing minimiser was applied to obtain these parameters. Errors for Univariate were found by assessing the curvature and errors for Simulated Annealing assessing NLL shifted by $0.5$. For Univariate, NLL $= 605.1$ and for Simulated Annealing, NLL $= 113.2$ where $\theta_{23} = 0.7278 \pm 0.0168 rad$, $\Delta m^2_{23} = (2.8541 \pm 0.0398) \times 10^3 eV^2$ and $\alpha = 1.6162 \pm 0.0803 m^2 eV^{-1}$.**

**Additionally, the methods were verified and tested using the Ackley and sphere function, with colour maps plotted to assess the steps taken by each minimiser. Error using the second derivative of the negative log-likelihood was also explored to assess closeness of fit. A 'Probabilistic Cooling Scheme' in Simulated Annealing was implemented to reduce temperatures efficiently.**

## I. Introduction

LIKELIHOOD represents a dimensional surface for N-1 parameters where by its maximum value describes the most suitable combination of parameters to fit a given sample of data [1]. It serves as a function of these parameters, formed by the product of their joint probabilities, given a sample of data [2]. In essence, it allows you to estimate a set of parameters for an equation to fit a given distribution of data.

For computational ease, the log of the likelihood is used for minimisation so that only sums rather than products are calculated. This does not affect the optimum values for the parameters you obtain. In this case, the negative of the log-likelihood was also taken so that the function could be minimised, instead of maximised. The curvature of the negative log-likelihood (NLL) indicates the stability for the parameter estimates, which provides a means to calculate the uncertainties on the estimate parameters in this project.

The Super-Kamiokande detector was the first to present this phenomenon of neutrino mixing in 1998, indicating that these neutrino particles had mass [3]. The three neutrino flavours are the: electron, muon and tau neutrinos and the mixing occurs due to a mixing of their mass and flavour eigenstates [4].

Neutrino oscillations are a statistically limited phenomena, requiring extremely sensitive equipment. Due to the nature of the low counts, the system can be described by a Poisson distribution if we consider the counts occuring in discrete energy bins. The NLL of a Poisson distribution is governed as

$$\mathrm{NLL}(\mathbf{u}) = \sum_{i=1}^{n} \left[ \lambda_i(\mathbf{u}) - k_i + k_i \log\left(\frac{k_i}{\lambda_i(\mathbf{u})}\right)\right] \quad (1)$$

where $\lambda_i$ is the predicted rates after oscillation, $k_i$ were the observed or measured rates after oscillation and $\mathbf{u}$ reflect a set parameters, in this project these were the mixing angle $\theta_{23}$, difference between neutrino masses squared $\Delta m^2_{23}$ and cross section rate $\alpha$. A point to note, if the $k_i$ of a certain energy is zero, the log component was set to zero as $\lim_{x\to 0}[x \log x] = 0$.

Survival probability determines how likely a neutrino is to remain as its current flavour, instead of changing its flavour. Given the sample of data, the survival probability for a given energy is found from the following equation:

$$P\left(v_\mu \longrightarrow v_\mu\right) = \left[1 - \sin^2\left(2\theta_{23}\right)\sin^2\left(\frac{1.267\Delta m^2_{23} L}{E}\right)\right]\alpha E \quad (2)$$

where $P$ is the probability $\mu$ represent a neutrino type, L is a fixed distance taken as $295$km, $E$ is the neutrino energy values in GeV and $\alpha$ is the cross section rate. When not considering the cross section, the $\alpha E$ factor is ignored and is set to one.

The predicted oscillated rates from the simulated unoscillated rates were calculated by finding the product between the survival probability and the unoscillated rates, this represented the value of $\lambda_i$ found in Eqn. 1.

From this, Eqn. 2 is substituted into the NLL to be minimised where the neutrino oscillation parameters are extracted. This process of minimising the negative log-likelihood function, given a sample of data and the survival probability for neutrino oscillations, represents the aim for this project.

## II. Theory

Two minimisers were tested in this project: Univariate and Simulated Annealing. Univariate is based on the parabolic method, and Simulated Annealing utilises features of thermodynamics to find the minimum.

### A. Univariate Parabolic Method

Consider the parabolic one dimensional minimum search method, where the function $f(x)$ is minimised. The method starts by taking a guess of three points $x_0$, $x_1$, $x_2$, where a 2nd order Lagrange polynomial is fitted according to these points. This Lagrange polynomial is just a parabola. The minimum of this parabola $x_3$ is then found by solving the first derivative set to zero, it is seen as:

$$x_3 = \frac{1}{2}\frac{\left(x_2^2 - x_1^2\right)y_0 + \left(x_0^2 - x_2^2\right)y_1 + \left(x_1^2 - x_0^2\right)y_2}{\left(x_2 - x_1\right)y_0 + \left(x_0 - x_2\right)y_1 + \left(x_1 - x_0\right)y_2} \quad (3)$$

where $x_{0,1,2}$ are the starting points and $y_{0,1,2}$ are the starting points evaluated using $f(x)$ respectively [6]. The values $f(x_{0,1,2,3})$ are evaluated and the three values with the

lowest $f(x_n)$ values are kept, with the last one discarded. The algorithm is then repeated for the new set of three points, progressing down the function until it converges.

Initial guess ranges are important when using the parabolic method. If the initial range includes a mixture of positive and negative curvatures, this could lead to Eqn. 3 finding the maximas of parabolas instead of minimas. To combat this, if $f(x_3)$ was found to be larger than all $f(x_{0,1,2})$ indicating a maxima, two additional points were generated in this range and evaluated, where the points with the three lowest $f(x_n)$ value were kept. This case indicates negative curvature where the minimiser begins to find a maxima. A warning was coded to inform the user.

### *B. Simulated Annealing*

Simulated Annealing is a probabilistic minimisation method based on the principle of thermodynamics. It is governed by the Boltzmann probability distribution, where random fluctuations could push the system into a states of higher energy, corresponding to a local minima. This allows the procedure to escape from local minima and search for the global minimum. The PDF is seen as

$$P(E)dE \sim \exp\left(-\frac{E}{k_B T}\right) dE \tag{4}$$

where $P$ is probability, $E$ is energy, $T$ is temperature and $k_B$ is a constant set to one. $E$ is replaced with a functional equation, which for this case is the negative log-likelihood Eqn. 1.

Points are accepted with the probability of:

$$p_{\mathrm{acc}} = \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp\left(-\Delta E / k_B T\right) & \text{if } \Delta E > 0 \end{cases} \tag{5}$$

where the $p_{acc}$ is the acceptance probability. The conditions for $p_{acc}$ indicate two types of steps. If $\Delta E \leq 0$, the subsequent step is a step in the correct direction to reduce 'energy' and is thus accepted 100% of the time. Otherwise if $\Delta E > 0$, this indicates a step that increases energy which is accepted with a probability, allowing the minimiser to escape local minima. This incorporates the random fluctuations characteristic. To accept it with a probability, this value was only accepted if it was larger than a uniformly randomly generated number between 0 and 1. If both cases reject the proposed point, a new point is then generated and the process is repeated until temperature reaches zero. Proposal points are generated using a gaussian distribution so that they are close to the previous point.

In essence, the algorithm performs a very broad and general search across all minimas in the beginning at high temperatures, where its probabilistic nature allows it to escape local minimas. From Eqn. 4, at large $T$, this causes a larger $p_{acc}$, increasing the chance to escape local minima. On the other hand, as $T \to 0$, $p_{acc} \to 0$ allowing the minimiser to settle at global minimum.

To increase the efficiency for this Simulated Annealing, a 'Probabilistic Cooling Scheme' was used to reduce the temperature using:

$$\mathrm{Ti} = \{\mathrm{A}/(\mathrm{i}+1) + \mathrm{B}\}\mathrm{PE} + \{\beta \text{ To } / \ln(1+\mathrm{i})\}\mathrm{PL} \tag{6}$$

$$\mathrm{A} = (\mathrm{T_o} - \mathrm{T_f})\,(\mathrm{N}+1)/\mathrm{N}; \mathrm{B} = \mathrm{T_o} - \mathrm{A} \tag{7}$$

where $N$ is the total number of steps, $i$ is the step number, $\mathrm{T_i}$ is the temperature at step $i$, $\mathrm{T_o}$ is this starting temperature set to 1000K, $\mathrm{T_f}$ is the final temperature, $\beta$ is the cooling factor that controls the increment size between temperatures and $PE$ and $PL$ are constants to control the cooling rates. For this project, $PE = 0.3$ and $PL = 0.29$ for optimum performance. $\beta$ was set to $0.01$ for efficiency [7].

The function contains an exponential and log component that combine to cool the temperature at different rates depending on its value.

For comparison, Simulated Annealing requires a longer time but can have a large search range before always converging on the global minima of the system. Univariate is much quicker and direct in its search since it is not probabilistic, allowing results to be reproducible. However it can get trapped in local minima and thus benefits from a small starting search range.

### *C. Errors*

Three types of error were found for NLL: linear, curvature and second derivative error.

The NLL value that was $+\frac{1}{2}$ greater than the minimum was used to estimate the error of the measurement, as propagation of errors was not available and the function was not approximately a Gaussian. Taking the Taylor Expansion of the log-likelihood to second order, evaluated at the minimum, gives

$$\begin{aligned} \ln[L(\theta)] &= \ln[L(\theta_m)] - \frac{(\theta - \theta_m)^2}{2\Sigma^2} \\ &= \ln[L(\theta_m)] - \frac{(\theta_m \pm \Sigma - \theta_m)^2}{2\Sigma^2} \\ &= \ln[L(\theta_m)] - \frac{1}{2} \end{aligned} \tag{8}$$

where $\theta$ is the parameter, $\theta_m$ is the maximum value, $\Sigma$ is the square root of the second derivative evaluated at the maximum and $\theta = \theta_m \pm \Sigma$. Since the negative log-likelihood was used, the values were taken at $+\frac{1}{2}$ instead of $-\frac{1}{2}$.

For linear error, NLL $+\frac{1}{2}$ was found through 1D linear interpolation. This provided two errors, on the positive and negative side of the minimum. Applying this to higher dimensions involved changing NLL w.r.t one parameter whilst keeping other parameters at their constant. The larger error was quoted in the abstract for formatting purposes only.

For the error in curvature, this value was found by fitting a Lagrange polynomial to the three lowest points generated by the parabolic minimiser. Since the fit equation had the form $Ax^2 + Bx + C$ where $Bx + C$ are just translation factors, the equation was solved for x by using

$$Ax^2 = +\frac{1}{2} \tag{9}$$

where A is the curvature, and $\frac{1}{2}$ is from the NLL error. This provided one value for uncertainty,

For the second derivative method, the difference between the curvatures $A$ from the NLL equation and the Lagrange polynomials were found to assess the closeness of fit. The second derivative of NLL is seen as

$$\frac{\partial^2 NLL}{\partial x^2} = \sum_{i=1}^{n} \left[\lambda'' - \frac{k_i(\lambda\lambda'' - \lambda'^2)}{\lambda^2}\right] \tag{10}$$

where $x$ is an arbitrary parameter and $'$ represents a derivative. This was only done w.r.t to $\theta_{23}$ as it only serves as a validation.

The curvature method provided a more precise uncertainty for the parabolic minimiser as it dealt specifically with parabolas. The linear fit requires less paramters for implementation, but depends on the accuracy used between points when interpolated. Although with a sufficiently high accuracy, the difference between methods would be small (at the cost of computation time). Both came out to give similar results within 1% of each other. Curvature was used for Univariate and linear method was used for Simulated Annealing.

### D. Convergence

To determine convergence in both methods, if the final 3 minimas generated were the same to 15 decimal points, the method was seen to have converged. 15 decimal points were used as the machine epsilon on the chosen hardware was to the order of $10^{-16}$.

### E. Test functions

Two test functions were used to verify the minimisers, the sphere function and the Ackley function.

The sphere function is seen as

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} x_i^2 \tag{11}$$

where the minimum lies at $f(x_0, x_1, ..., x_n) = 0$.

The Ackley function is seen as

$$\begin{aligned} f(x,y) = &-20\exp[-0.2\sqrt{0.5\,(x^2+y^2)}] \\ &- \exp[0.5(\cos 2\pi x + \cos 2\pi y)] + e + 20 \end{aligned} \tag{12}$$

where the minimum lies at $f(0,0) = 0$ for a search range of $[-5, 5]$.

## III. Results and Discussion

### A. Results

Parameters from the 2D Univariate method and the 3D Simulated Annealing can be seen in the Tables I and II respectively. For Simulated Annealing, the starting temperature of $1000K$ was used, with more than 20,000 steps taken following the 'Probabilitistic Cooling Scheme', before ending at $0K$. It should be noted that different combinations of the parameters will still provide the same minimum NLL, and this is one combination.

From data exploration, the simulated unoscillated rates and measured oscillated rates can be seen in Fig. 1. Upon applying survival probability to the unoscillated rates, it is expected that the new predicted oscillated rates follow the same distribution as the measured data.

TABLE I
Univariate with Curvature Error

| Parameter | Value | Error |
|---|---|---|
| $\theta_{23}$ | 0.7121 | $\pm 0.0144$ |
| $\Delta m_{23}$ | 0.0027631 | $\pm 0.0000324$ |
| NLL(**u**) | 605.1 | |

TABLE II
Simulated Annealing with Linear Error

| Parameter | Value | Lower, Upper Error |
|---|---|---|
| $\theta_{23}$ | 0.7278 | $-0.0168, 0.0118$ |
| $\Delta m_{23}$ | 0.0028541 | $-2.31 \times 10^5, 3.98 \times 10^5$ |
| $\alpha$ | 1.6162 | $-0.0456, 0.0803$ |
| NLL(**u**) | 113.2 | |

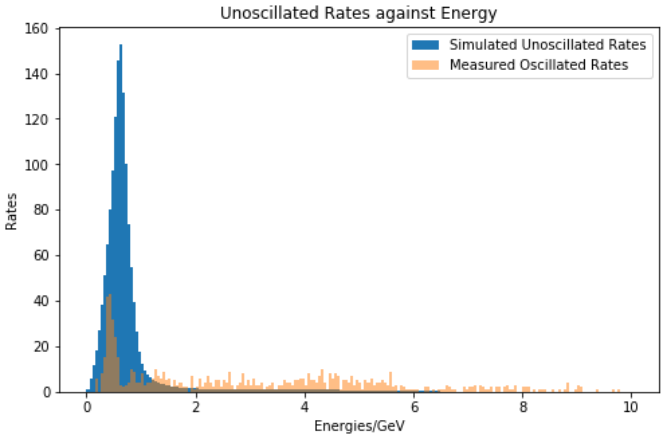


Fig. 1. Plot of the measured oscillated rates against simulated unoscillated rates. The simulated data is expected to follow the same distribution as the measured rates after applying the survival probability.

Cross section was assumed to be linearly proportional to energy, following

$$\sigma = \alpha E + C \tag{13}$$

where $\sigma$ is cross section and $\alpha$ and $C$ are constants. The NLL value was seen to reduce by 81% when comparing between the 2D and 3D case. This enforces the idea that the cross section is a suitable parameter to neutrino oscillations. $C$ was set to zero in this project, this was verified using a 4D Simulated Annealing minimisation, yielding results that $C$ equaled to zero to 2 significant figures.

When applying the Univariate method, $\Delta m_{23}$ was minimised before minimising $\theta_{23}$. This is because NLL against $\Delta m_{23}$ displayed many more local minima when $\theta_{23}$ was near $\frac{\pi}{4}$. Setting a low initial guess for $\theta_{23}$ (approx. 0.1), produces a smoother function for NLL against $\Delta m_{23}$, reducing the chances of getting stuck in a local minima, increasing the speed of the method.

As the value of $\theta_{23}$ tends to $\frac{\pi}{4}$, the bump seen at $\frac{\pi}{4}$ reduces and $\theta_{23}$ tends to a minimum value, and hence errors at this point would be symmetrical. The bump does always reduce as $\theta_{23} \to \frac{\pi}{4}$, avoiding the incorrect case where points are

interpolated in a manner that neglects the local minima.

Simulated Annealing accepted points at approximately 0.18% out of 24027 steps.

### B. Validations

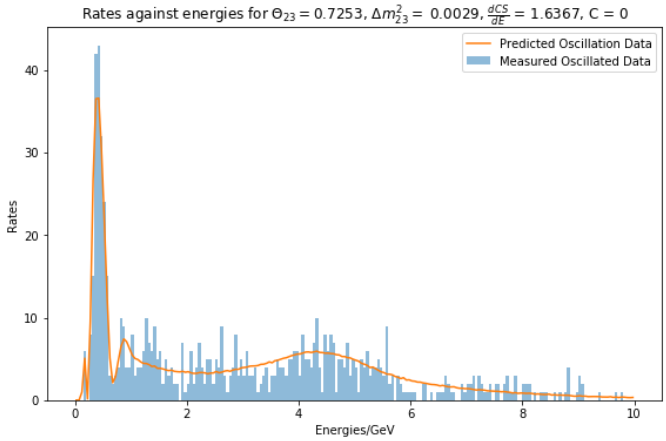


Fig. 2. Plot of rates against energies, using the model parameters in Table II from 3D minmisation. The predicted oscillation data can be seen to match the measured data, except for the peak at low energies. This could be matched by introducing and minmising a new parameter.

Values from Table II produced the following rates against energy plot in Fig. 2. The predicted oscillation rates matches the measured data across the whole distribution except for the peak at low energies. This indicates that NLL was minimised to denote parameters that are representative of the physical phenomena, but could be further improved by possibly miniminising an additional parameter. Ideally, the predicted data should match the measured data exactly.

Parabolic minimisations were validated using the second derivative method described by Eqn. 10. For 1D Parabolic, $\Delta A = 0.025$ and for 2D Univariate, $\Delta A = 0.291$. These values show good curvature fits to the original function, thus accurate minimisations.

The accepted steps for simulated annealing was plot over a 2D colour map to assess the effectiveness of the method in finding global minimas.

For Univariate, 1D steps that progressively minimise the value for NLL can be seen in Fig. 3, verifying that the Univariate method is working as expected.

For the Simulated Annealing, these steps are seen in Fig. 4. This is seen to converge towards the global minima in a few steps, before jumping between both local minimas until it converging on one. The function is symmetric about $\frac{\pi}{4}$, thus both values about the symmetry are accepted as model values.

The sphere function where the value of x is offset by 1 and was used to test Univariate method. This returned the expected minimas exactly. The Ackley function was tested for both methods. For the Univariate method, the exact expected parameter values were found. For Simulated Annealing, the accepted steps using the Ackley function can be seen in Fig. 5, with its parameter values seen in Table III.

It can be seen from Fig. 5, that Simulated Annealing performs a broad general search across many local minimas,

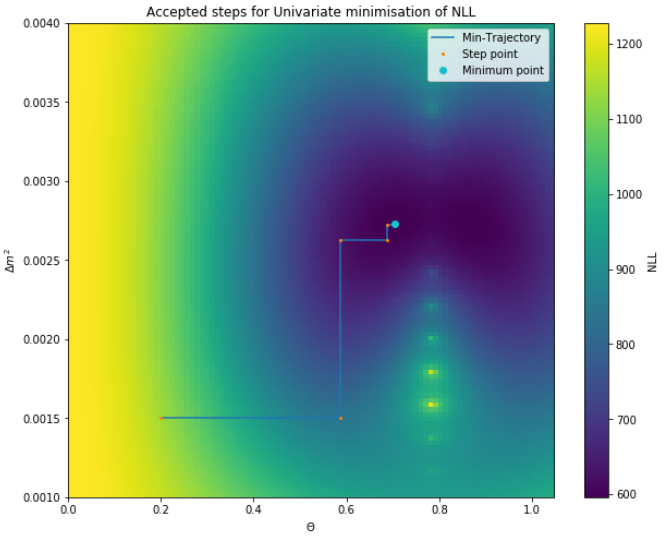


Fig. 3. Plot of the accepted steps from 2D minimisation using the Univariate method overlaid on a 3D colour map. Orange points represents an accepted step, the blue lines are the trajectory to the next accepted step, and the cyan point represents the value that minimises NLL

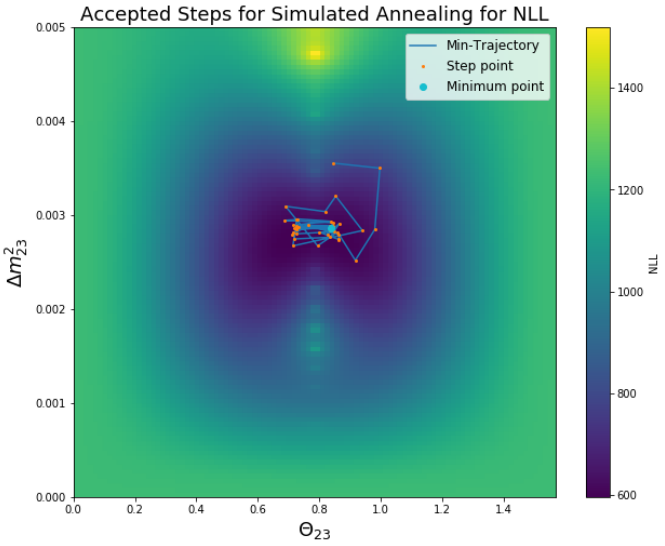


Fig. 4. Plot of the accepted steps from 3D minimisation using the Simulated Annealing overlaid on a 3D colour map. Orange points represents an accepted step, the blue lines are the trajectory to the next accepted step, and the cyan point represents the value that minimises NLL

occasionally jumping to higher energy states, before converging on the global minimum. This validation returned the global minimum of the Ackley function each time it was run.

TABLE III
SIMULATED ANNEALING FOR ACKLEY

| Parameter | Measured | Expected |
|---|---|---|
| $x$ | $-0.00356$ | 0 |
| $y$ | 0.00028 | 0 |
| $f(x, y)$ | 0.01043 | 0 |

Assessing Table III, the parameter values are within 1% of the expected parameter values. It is possible to obtain more

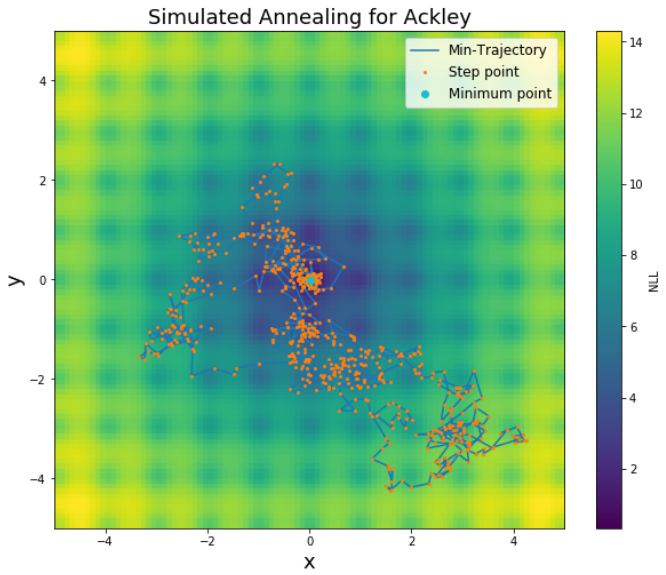


Fig. 5. Plot of the accepted steps from 3D minimisation using the Simulated Annealing for the test Ackley function, overlaid on a 3D colour map. Orange points represents an accepted step, the blue lines are the trajectory to the next accepted step, and the cyan point represents the value that minimises NLL

accurate values by increasing $\beta$ in Eqn. 6 to evaluate it over a larger number of steps. However, this comes at the cost of time. Using the current parameters stated in section II: Theory, the run-time approximates to 20 seconds for computation and hence is still reasonable for experimental testing. As $\beta \to 1$, time taken tends to infinity.

## IV. Conclusion

The aim of the project was to investigate the model parameters to assess the survival probability describing neutrino oscillations, by minimising the negative log-likelihood. This was achieved by utilising a Univariate Parabolic and Simulated Annealing minimiser.

From this project, an understanding for the utility of minimisation in physical discovery has been achieved, as well as nuanced understanding of the merits of both minimisers. Much time was spent on exploring the physical interpretation of this project, inspiring insight into neutrino physics.

Successes would be that the method was able to draw from NLL to produce parameters that resemble the measured data. 3D plots could have been produced to better portray the problem and additional minimisers could be tested for experimentation.

## References

[1] Lee, Jeffrey (2009). "Curves and Hypersurfaces in Euclidean Space". Manifolds and Differential Geometry. Providence: American Mathematical Society. pp. 143–188.

[2] Myung, In Jae (2003). "Tutorial on Maximum Likelihood Estimation". Journal of Mathematical Psychology.

[3] Fukuda,Y, et al. "Evidence for oscillation of atmospheric neutrinos" (1998). Physical Review Letters.

[4] Fukuda, Y.; Hayakawa, T.; Ichihara, E.; Inoue, K.; Ishihara, K.; Ishino, H.; et al. (1998). "Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days". Physical Review Letters.

[5] Ackley, D. H. (1987) "A connectionist machine for genetic hillclimbing", Kluwer Academic Publishers, Boston MA

[6] Press.W, Teukolsky.S, Vetterling.W, Flannery.B (2007). "Numerical Recipes 3rd Edition: The Art of Scientific Computing", Cambridge University Press.

[7] Samuel.R, Venkumar.P (2015). "Optimized Temperature Reduction Schedule for Simulated Annealing Algorithm", Materials Today: Proceedings.