Minimisation of a negative log likelihood fit to extract the lifetime of D^0 meson

Son-Gyo Jung (CID: 00948246)
Department of Physics, Imperial College London
(Dated: December 19, 2016)

Using the Python programming language, the average lifetime of D^0 meson was extracted from 10,000 experimental data of decay time and its error by minimising the negative log-likelihood (NLL) corresponding to cases with and without the background signals. In the absence of possible background signals, the parabolic minimisation method was employed, yielding the average lifetime as $(404.5 \pm 4.7) \times 10^{-15}$s with a tolerance level of 10^{-6}. This result was found to be inconsistent with the literature value provided by the Particle Data Group, showing a deviation of $\sim 6 \times 10^{-15}$s. By considering possible background signals, an alternative distribution and the corresponding NLL were derived. This was subsequently minimised using the gradient, Newton's and the Quasi-Newton methods, yielding consistent results. The average lifetime and the fraction of the background signals in the sample were estimated to be $(409.7 \pm 5.5) \times 10^{-15}$s and 0.0163 ± 0.0086, respectively, where the uncertainties were calculated using an error matrix and the correlation coefficient was found to be -0.4813. The literature value lies within the uncertainty, showing a percentage difference of $\sim 0.098\%$. Thus the results verify the presence of the background signals in the data and validate the theory of the expected distribution derived by assuming the background signal as a Gaussian due the limitation of the detector resolution.

1. Introduction

Most subatomic particles are unstable with a typical lifetime τ of the order of 10^{-12} seconds and consequently a direct measurement of τ is impossible. However, measuring the momentum of the individual particles and the corresponding vector, which connects the production and the decay point, provides an indirect means of estimating τ. In this analysis, it is of particular interest to computationally extract the average lifetime of D^0 meson and its associated uncertainty by minimising the negative log likelihood (NLL) function with and without possible background signals.

For a given tolerance level, the initial minimisation of NLL function involved utilising the parabolic method, from which the lifetime corresponding to the minimum τ_{min} was calculated. The uncertainties in τ_{min} were subsequently estimated using the NLL function as well as the curvature of the last parabolic estimate. This was achieved by employing the bisection method and the secant method to approximate τ^+ and τ^-, corresponding to values of τ when the minimum of the NLL or the parabola are increased by 0.5.

Following this, the possibility of background signals in the dataset was considered. The lifetime and the fraction of the signal in the sample a, and thus the fraction of background signals, were estimated by minimising an alternative NLL function that depended on both τ and a. The minimisation was carried out using the gradient, Newton's and the Quasi-Newton methods, while the uncertainties in τ_{min} and a_{min} were likewise estimated from the approximation of τ^+ and τ^-. In addition, the error matrix, which considers possible correlation between τ and a, was also studied.

2. Theory

2.1 Expected distribution of decay time

As mentioned in Section 1, an indirect approach is required to measure the lifetime of subatomic particles. From relativistic dynamics, the decay time t can be obtained from the momentum $\mathbf{p}$ of the individual particles and the corresponding production - decay vector $\mathbf{x}$ via:[1]

$$t = m\frac{\mathbf{x}\cdot\mathbf{p}}{p^2}, \tag{1}$$

where m is the mass of the particle. Following a significant amount of decays of particles that are of the same type, the expected decay times (Eq. (1)) will follow the exponential distribution $f^t(t)$ given by:[1]

$$f^t(t) = \begin{cases} 0 & \text{if } t < 0 \\ \frac{1}{\tau}\exp(-\frac{t}{\tau}) & \text{if } t \geq 0 \end{cases} \tag{2}$$

where τ is the average lifetime that is to be determined. As a result of the uncertainty associated with the production and decay spatial measurements, however, the exponential distribution (Eq. (2)) becomes no longer adequate for modelling the dataset and each measurement of t will have an associated uncertainty σ. Therefore, by taking the uncertainty into account, the expected distribution of the measured t is given by a convolution of Eq. (2) with a Gaussian of width σ as shown below:[1]

$$\begin{aligned} f^m(t) &= f^t(t) \otimes G(\sigma) \\ &= \int_{-\infty}^{\infty} \frac{1}{\tau}\exp\left(\frac{-t'}{\tau}\right)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(t-t')^2}{\sigma^2}\right)\mathrm{d}t' \\ &= \frac{1}{2\tau}\exp\left(\frac{\sigma^2}{2\tau^2} - \frac{t}{\tau}\right)\mathrm{erfc}\left(\frac{1}{\sqrt{2}}(\frac{\sigma}{\tau} - \frac{t}{\sigma})\right), \end{aligned} \tag{3}$$

where erfc is the error function and σ is dependent on the kinematics of the decay, which varies from one measurement to another. Thus Eq. (3), which is normalised, acts as a probability density function that can be utilised in a statistical analysis for the case where background signals are ignored or negligible.[1]

2.2 Expected distribution with background signals

In practice, there are small background signals or noises in the dataset that are random in nature and have zero lifetime. However, due the limitation of the resolution that can be achieved by a detector, these random noises undergo a smearing effect in a similar manner to the data from an actual decay, misleading one to think that the background noises correspond to real decay events. Therefore, by invoking the central limit theorem, the background can be assumed to be a convolution of a delta function $\delta(t)$ with a Gaussian, which is itself a Gaussian.

This is expressed as:[1]

$$\begin{aligned} f^m_{bkg}(t) &= \delta(t) \otimes G(\sigma) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{t^2}{\sigma^2}\right) \\ &= G(\sigma), \end{aligned} \tag{4}$$

with the symbols defined in Section 2.1.

With the distribution of the background defined, a more realistic distribution of t can be defined as a function $f^m_{tot}(t)$ that is dependent on both τ and the fraction of the signal in the data a as shown below:[1]

$$f^m_{tot}(t) = a f^m_{sig}(t) + (1-a) f^m_{bkg}(t), \tag{5}$$

where $f^m_{sig}(t)$ is the expected distribution corresponding to the true decay events (Eq. (3)). Therefore, Eq. (5) is the probability density function in the presence of background signals.

2.3 The average lifetime of D^0 meson

A typical high energy physics experiment yields significantly large number of data and thus statistical approaches are taken in order to analyse the data. In this analysis, 10,000 pairs of $(t,\ \sigma)$ are utilised. Since σ depends on the kinematics of the decay, causing the value to vary from one measurement to another, the average τ cannot be extracted through the use of a traditional χ^2 fit to a binned distribution. The alternative approach is to use a negative log-likelihood (NLL) fit to estimate the parameter τ (and a when considering the background signals). The theory associated with the NLL fit is as follows.

For a given n independent data points with measurement $\mathbf{m}_i$ and a probability density function $\mathcal{P}$, the total likelihood $\mathcal{L}$ is the product of the separate probabilities, which is expressed as:[1,2,3]

$$\mathcal{L} = \prod_{i=1}^{n} \mathcal{P}(\mathbf{m}_i), \tag{6}$$

where $\mathcal{P}(\mathbf{m}_i)$ is the likelihood of a given measurement $\mathbf{m}_i$. Eq. (6) is valid as long as the measurements are uncorrelated. Since the measurements $\mathbf{m}_i$ are fixed, if the probability density function depends on an unknown parameter(s) $\mathbf{u}$, the total likelihood is purely a function of $\mathbf{u}$. This can be written as:[1,2,3]

$$\mathcal{L}(\mathbf{u}) = \prod_{i=1}^{n} \mathcal{P}(\mathbf{u};\ \mathbf{m}_i). \tag{7}$$

Eq. (7) is simply stating that the likelihood is effectively the probability of seeing the data values actually observed as a function of the parameter.

In practice, logarithms are taken on both sides of Eq. (7) as the logarithm is monotonically increasing function, yielding a log-likelihood. This means the maximum of $\mathcal{L}(\mathbf{u})$ is also the maximum of $\log(\mathcal{L}(\mathbf{u}))$. Therefore, the value of the parameter $\mathbf{u}_m$ that maximises the the likelihood function will simultaneously maximise the log-likelihood function, and it will correspond to the best fit between the dataset and the probability density function. Subsequently, $\mathbf{u}_m$ can be taken as an estimator of the true value of $\mathbf{u}$.

Similarly, the minimisation of the negative log-likelihood (NLL) can be carried out instead of maximising $\log(\mathcal{L}(\mathbf{u}))$ to calculate the estimator of the true value of $\mathbf{u}$. The NLL function is given by:

$$\begin{aligned} \mathrm{NLL}(\mathbf{u}) &= -\log\left(\prod_{i=1}^{n} \mathcal{P}(\mathbf{u};\ \mathbf{m}_i)\right) \\ &= -\sum_{i=1}^{n} \log(\mathcal{P}(\mathbf{u};\ \mathbf{m}_i), \end{aligned} \tag{8}$$

where the concept is analogous to the positive log-likelihood.

3. Method

3.1 Preliminary approximation of the lifetime

The distribution of the experimental data of (t_i, σ_i), both in units of picoseconds, was graphically analysed by creating normalised histograms of various bin sizes, allowing a rough profile of the probability distribution to be inferred. Subsequently, the fit function Eq. (3) with different values of τ and σ were plotted together with the histogram from which a preliminary approximation of τ was obtained. This was achieved by identifying the corresponding value of τ of a fit function that was perceived to depict the profile of the histogram the most accurately.

In order to validate the preliminary approximation of τ, the normalisation of Eq. (3) was verified computationally by integrating the function over t and demonstrating that it is approximately equal to unity with a negligible error. Furthermore, the independence of the integral of Eq. (3) from the values of τ and σ was also illustrated, confirming the validity of the preliminary result.

3.2 Minimisation of NLL function

As explained in Section 2.3, the value of the parameter $\mathbf{u}_m$ that minimises the NLL function can be taken as an estimator of the true value of $\mathbf{u}$. Therefore, in order to extract the average lifetime of the D^0 meson from the dataset, the NLL function was minimised using various computational methods. In this analysis, the notations in Eq. (8) will be defined as follows:

†Without background signals

$$\mathcal{P}(\mathbf{u};\ \mathbf{m}_i) = f^m(t) = f^m_{sig}(t), \tag{9a}$$

where

$$\mathbf{m}_i = (t_i, \sigma_i), \tag{9b}$$

$$\mathbf{u} = \tau. \tag{9c}$$

‡With background signals

$$\mathcal{P}(\mathbf{u};\ \mathbf{m}_i) = f^m_{tot}(t), \tag{10a}$$

where

$$\mathbf{m}_i = (t_i, \sigma_i), \tag{10b}$$

$$\mathbf{u} = (\tau,\ a). \tag{10c}$$

With the definitions above, iterative methods were employed for the minimisation of the NLL function; with† and without‡ the possible background signals.

In the *absence*† of the background signal, there is only one unknown parameter; that is τ. Hence, extracting the average lifetime becomes a one-dimensional minimisation problem. By examining the graph of the NLL as a function of τ, it was evident that the parabolic method was suitable as the function closely approximates a parabola near the minimum. The parabolic method is as follows:[2,3]

†Parabolic method

For a function $f(x)$, where the curvature is positive within a given interval, three points x_0, x_1 and x_2 along $f(x)$ are chosen. The second-order Lagrange polynomial $P_2(x)$ is then fitted through the points, allowing quadratic interpolation to be made where $P_2(x)$ is given by:[2,3]

$$P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}y_2, \tag{11}$$

where $f(x_0) = y_0$, $f(x_1) = y_1$ and $f(x_2) = y_2$. Following this, the position of the minimum of the interpolating parabola $P_2(x)$ is calculated by differentiating $P_2(x)$ and equating it to zero. Thus, solving $\frac{\mathrm{d}P_2(x)}{\mathrm{d}x} = 0$ for x yields an estimate of the minimum x_3 as:

$$x_3 = \frac{1}{2}\frac{(x_2^2 - x_1^2)y_0 + (x_0^2 - x_2^2)y_1 + (x_1^2 - x_0^2)y_2}{(x_2 - x_1)y_0 + (x_0 - x_2)y_1 + (x_1 - x_0)y_2}. \tag{12}$$

Subsequent to identifying the new minimum, the highest point from $f(x_0)$, $f(x_1)$ and $f(x_2)$ is replaced with $f(x_3)$, and the above procedure is iterated so that x_3 converges towards the true minimum of $f(x)$. The iteration is terminated once Δx_3 becomes smaller than a desired precision (tolerance level). This method was employed for the 1D minimisation problem with $f(x) = \mathrm{NLL}(\tau)$.

In the *presence*‡ of the background signal, there are two unknown parameters; τ and a. Now, extracting the average lifetime becomes a 2D minimisation problem. The 1D parabolic method can be extended for higher dimensions (univariate method) in order to extract the average values of τ and a, searching for the minimum in each direction successively and iterating as mentioned before. The rate of convergence, however, is slow as the univariate search spirals into the minimum. Therefore, alternative methods, namely the gradient, Newton's and the Quasi-Newton methods were employed. The principle of these methods are as follows:[2]

‡Gradient method

The gradient method is a first-order iterative optimisation algorithm which involves following the steepest descent towards the minimum by evaluating the local gradient at each iteration. This method can be generalised as:

$$\vec{x}_{n+1} = \vec{x}_n - \alpha\vec{\nabla}f(\vec{x}_n), \tag{13a}$$

where

$$\vec{\nabla}f(\vec{x}_n) = \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_N \end{pmatrix}, \tag{13b}$$

and $\alpha \ll 1$ is the step-length that is positive. Eq. (13a) in effect is describing a movement perpendicular to the local contour line down towards the minimum as long as $f(\vec{x}_{n+1}) < f(\vec{x}_n)$ is satisfied.

‡Newton's method

More efficient minimisation methods utilise the local curvature (Hessian) at each iteration. Newton's method has a greater rate of convergence (quadratic convergence) than that of the gradient method in addition to being affine invariant. It requires calculating the Hessian $\mathbf{H}$, second partial derivatives of a function $f(\vec{x})$, given by:[2]

$$H_{ij}(\vec{x}) = \frac{\partial^2 f(\vec{x})}{\partial x_i \partial x_j}, \tag{14}$$

where i and j denote the row and column numbers of the Hessian matrix, respectively. In order to locate the minimum, $\mathbf{H}$ must be positive definite and satisfy:[2,3]

$$\mathbf{H}(\vec{x}_n)\cdot\vec{\delta}_n = -\vec{\nabla}f(\vec{x}_n), \tag{15}$$

where $\vec{\delta}_n = \vec{x}_{n+1} - \vec{x}_n$ is the update in $\vec{x}$, and Eq. (15) can be derived using the Taylor expansion to first order of $\vec{\delta}$. Thus, the position of the minimum can be estimated via the iteration step given by:[2]

$$\vec{x}_{n+1} = \vec{x}_n - [\mathbf{H}(\vec{x}_n)]^{-1}\cdot\vec{\nabla}f(\vec{x}_n), \tag{16}$$

where the exact position of the minimum can be obtain if $f(\vec{x})$ is parabolic.

‡Quasi-Newton method

An alternative method which utilises the local curvature at each step is the Quasi-Newton method. This method can be advantageous over Newton's method since the latter involves the Hessian which requires computing both the first and second derivatives of the multi-dimensional function; the inverse Hessian which in practice can cause the overall computational time to dramatically increase, taking $\mathcal{O}(N^3)$ operations per iteration. Moreover, calculating the Hessian using finite difference schemes often yields a matrix that is not strictly positive definite, becoming sensitive to the initial conditions and other parameters such as the step-size. The Quasi-Newton method avoids these problems by approximating the inverse Hessian $\mathbf{H}^{-1}(\vec{x}_n)$ using the local gradient. The basic iteration step is given by:[2]

$$\vec{x}_{n+1} = \vec{x}_n - \alpha\mathbf{G}_n\cdot\vec{\nabla}f(\vec{x}_n), \tag{17}$$

where $\mathbf{G}_n$ is an approximation of the inverse Hessian. For the initial iteration $\mathbf{G}_0$ is defined to be the identity matrix $\mathbf{I}$, making the first iteration identical to the gradient method. The vector $\mathbf{G}_n$ is subsequently updated to $\mathbf{G}_{n+1}$ via the Davidon-Fletcher-Power (DFP) algorithm, where the update is given by:[2]

$$\mathbf{G}_{n+1} = \mathbf{G}_n + \frac{(\vec{\delta}_n \otimes \vec{\delta}_n)}{\vec{\gamma}_n\cdot\vec{\delta}_n} - \frac{\mathbf{G}_n\cdot(\vec{\delta}_n\otimes\vec{\delta}_n)\cdot\mathbf{G}_n}{\vec{\gamma}_n\cdot\mathbf{G}_n\cdot\vec{\gamma}_n}, \tag{18}$$

where

$$\vec{\delta}_n = \vec{x}_{n+1} - \vec{x}_n = -\alpha\ \mathbf{G}_n\cdot\vec{\nabla}f(\vec{x}_n), \tag{19}$$

$$\vec{\gamma}_n = \vec{\nabla} f(\vec{x}_{n+1}) - \vec{\nabla} f(\vec{x}_n), \tag{20}$$

and $(\vec{\delta}_n \otimes \vec{\delta}_n)$ is the outer product of $\vec{\delta}_n$ and $\vec{\delta}_n$ while Eqs. (19) and (20) are the updates in $\vec{x}$ and $\vec{\nabla} f$, respectively. The DFP algorithm above allows the updated $\mathbf{G}$ to satisfy:[2]

$$\mathbf{G}_n \cdot \vec{\gamma}_n = \vec{\delta}_n. \tag{21}$$

Hence, Eq. (17) can be used to approximate the position of the minimum with $\mathcal{O}(N^k)$ operations at each iteration, where $k < 3$.

The minimisation methods discussed in this section for both with† and without‡ the background signals do not take into account the possibility of multiple minima in the interval of interest. In general, there is no fool-proof method to find the global minimum of a function. Therefore, the algorithms were applied from different starting points in order to validate the results.

Note: $\vec{\nabla} f(\vec{x}_n)$ was solved numerically using finite difference approximations, namely the forward difference scheme (FDS) and the central difference scheme (CDS), which are first and second-order accurate, respectively.

3.3 Error analysis

The accuracy of τ and a values are dependent on the quantity of numerical data available. In general, the uncertainty can be evaluated by changing the parameter and identifying the point which makes the log-likelihood change by $-\frac{1}{2}$ from the maximum in either positive or negative direction. This change corresponds to a shift of one standard deviation 1σ. Therefore, an increase in the NLL function by $\frac{1}{2}$ from the minimum in both directions were computed to calculate τ^+ and τ^-. The same procedure was carried out using the curvature of the last parabolic estimate, allowing a comparison of the uncertainties to be made which in turn validated the computational algorithms. Both methods are summarised below.

The estimate of the uncertainty using the NLL function involved solving the following equation:

$$\mathrm{NLL}(\tau) - \mathrm{NLL}(\tau_{min}) - 0.5 = 0, \tag{22}$$

while for the late parabolic estimate the equation solved was:

$$r(\tau - \tau_{min})^2 - 0.5 = 0, \tag{23}$$

where r is the curvature of the last parabolic estimate. In each case, the solution was estimated using iterative methods, where the initial guesses are refined iteratively. The methods were the bisection and the secant methods. The bisection method started with two initial guesses; one to the left x_L and one to the right x_R of the root. The average of the two guesses was calculated and the corresponding function value was evaluated. One of the initial guesses was subsequently updated with the average value depending on the polarity of the function. This was repeated until $|x_L - x_R| \leq \epsilon$, where $\epsilon = 10^{-6}$ is the tolerance level. This method converges linearly, with the uncertainty in the position of the root reducing by a factor of two with each iteration.

The secant method, in contrast, involves approximating the tangent to the function using two points on the curve and utilising the Newton-Raphson method, which gives:[2]

$$x_{i+1} = x_i - f(x_i)\frac{x_i - x_{i-1}}{f(x_i) - f(x_{i-1})}. \tag{24}$$

The speed of convergence of the secant method is somewhere between linear and quadratic. This method was used in order to validate the results obtained using the bisection method.

The methodologies in estimating the error discussed above is for the 1D problem. For the 2D problem, the methods are analogous with four distinct directions; two for τ and two for a. When the error in τ was analysed, the value of a was set to the value corresponding to the minimum of the NLL function. This was likewise true when analysing the error in the a value.

It is important to note that the error analysis in the 2D case assumes that the parameters are independent; i.e. uncorrelated. The general formulation for uncertainties with more than one parameter utilises the error matrix $\mathbf{E}$, which is also referred to as the covariance matrix. This is given by:[3]

$$\mathbf{E} = \begin{pmatrix} \sigma_\tau^2 & \rho_{\tau,a}\sigma_\tau\sigma_a \\ \rho_{\tau,a}\sigma_\tau\sigma_a & \sigma_a^2 \end{pmatrix}, \tag{25}$$

where $\rho_{\tau,a}$ is the correlation coefficient between τ and a. Eq. (25) can be obtained by using the relationship:[3]

$$\mathbf{W} = \mathbf{E}^{-1}, \tag{26}$$

where the elements in the weight matrix $\mathbf{W}$ are given by:[3]

$$W_{ij} = \left.\frac{\partial^2 \mathrm{NLL}(\tau, a)}{\partial\theta_i \partial\theta_j}\right|_{\hat{\theta} = \tau_{min} \text{ or } a_{min},} \tag{27}$$

where $\theta_i = \tau$ and $\theta_j = a$. Subsequently, Eqs. $(25 - 27)$ were used to study the correlation between τ and a in addition to estimating more plausible uncertainties taking the correlation between the parameters into account.

4. Results and Discussions

4.1 Without background signals

4.1.1 Distribution of the decay time

The distribution of the experimental decay time was initially analysed by creating a normalised histogram with 500 bins, and plotting the fit function with different values of τ and σ. This is shown in FIG. 1.

FIG. 1 suggests that the average life time τ is ~ 0.4ps while σ of 0.2ps appears to best match the overall spread. The fit function $f^m(t)$, which is the probability density function, was found to be independent of both τ and σ as expected. By integrating the fit function over the range $-5 < t < 10$, it was shown that the area of each fit was approximately equal to one with a negligible error. The figure also illustrates the possibility of τ being negative. Despite the theoretical distribution function being the exponential distribution $f^t(t)$ given by Eq. (2), where the $f^t(t) = 0$ when $t < 0$, the expected distribution is a convolution of the exponential distribution with a Gaussian

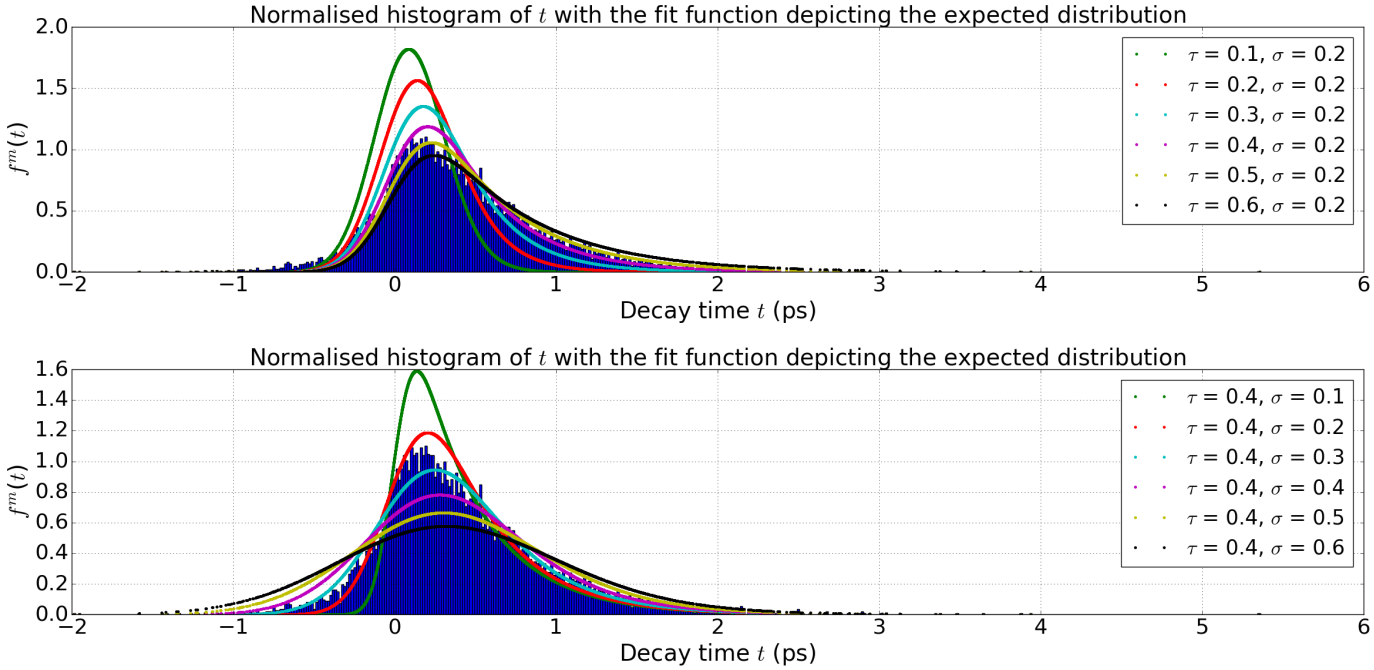

FIG. 1: Histogram of the measured decay times of D^0 meson and the expected distribution with various τ and σ in the units of picoseconds. The figure illustrates that the average lifetime is approximately between 0.4ps and 0.5ps, being closer to the former value. The second figure clearly demonstrates that $\tau = 0.4$ps with $\sigma = 0.2$ps depicts the profile of the histogram the most closely. Furthermore, integrating each fit yielded a result that was ~ 1 with an error of order 10^{-10}.

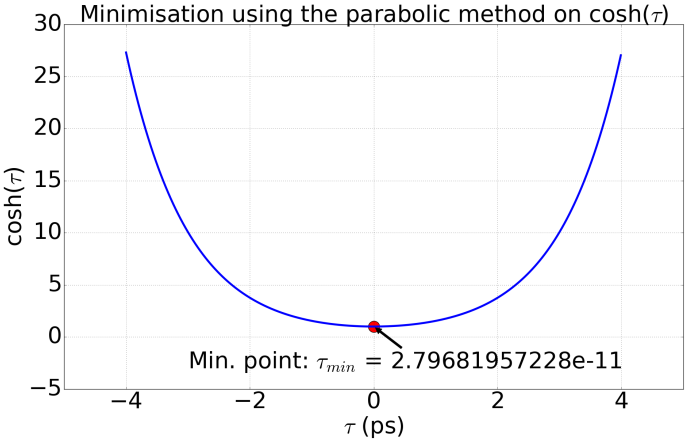

FIG. 2: Result of the minimisation using the parabolic method on a hyperbolic cosine function. The initial guesses were 2ps, 3ps and 5ps and the minimum is estimated to be at $\tau = 2.80 \times 10^{-11}$ (3 s.f.) using a tolerance level of 10^{-6}.

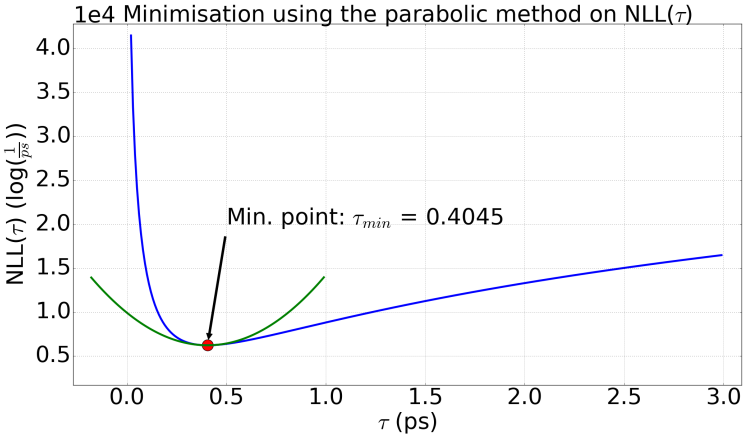

FIG. 3: Graph of the 1D NLL. The minimisation yielded the minimum as $\tau_{min} = 0.4045$ps correct to 4 d.p. with a tolerance level of 10^{-6}. The minimum was originally estimated to be roughly 0.40ps, which is equal to the result correct to 2 d.p. Moreover, the parabola with a curvature of 22572 illustrates its suitability in approximating the minimum.

due to the uncertainty associated with the spatial measurements of the production and decay points. Thus, the results are as expected and consistent with the theory.

Moreover, it should be noted that decreasing the value of τ shifts the maximum towards zero with the distribution roughly resembling a profile of a Gaussian. On the other hand, increasing the value of σ causes the distribution to broaden, becoming more symmetrical with the position of the maximum shifting slightly to the right.

4.1.2 Minimisation of the 1D NLL

To minimise the 1D NLL, a function was written and its functionality was tested using a hyperbolic cosine function. The result shown in FIG. 2 clearly demonstrates that the 1D minimisation using the parabolic method is adequate. Subsequent to minimising the NLL function, the position of the minimum was roughly approximated using a plot of the NLL function (FIG. 3). FIG. 3 also shows the parabola from which the value of τ_{min} and its uncertainty were calculated.

With the position of the minimum calculated, the uncertainty in τ_{min} was estimated as mentioned in Section 3.3. The results are in summarised TABLE I.

TABLE I: Summary of the uncertainties in τ_{min}

$\tau_{min} = 0.4045$ps	Bisection method		Secant method	
Function	σ^+ (ps)	σ^- (ps)	σ^+ (ps)	σ^- (ps)
Parabolic:	0.0047	0.0047	0.0047	0.0047
NLL:	0.0047	0.0047	0.0047	0.0047

The two different approaches, examining $|\Delta\text{NLL}| = \frac{1}{2}$ and using the curvature of the last parabolic estimate, yielded the uncertainties with equal magnitude correct to 4 d.p. with both the bisection and the secant meth-

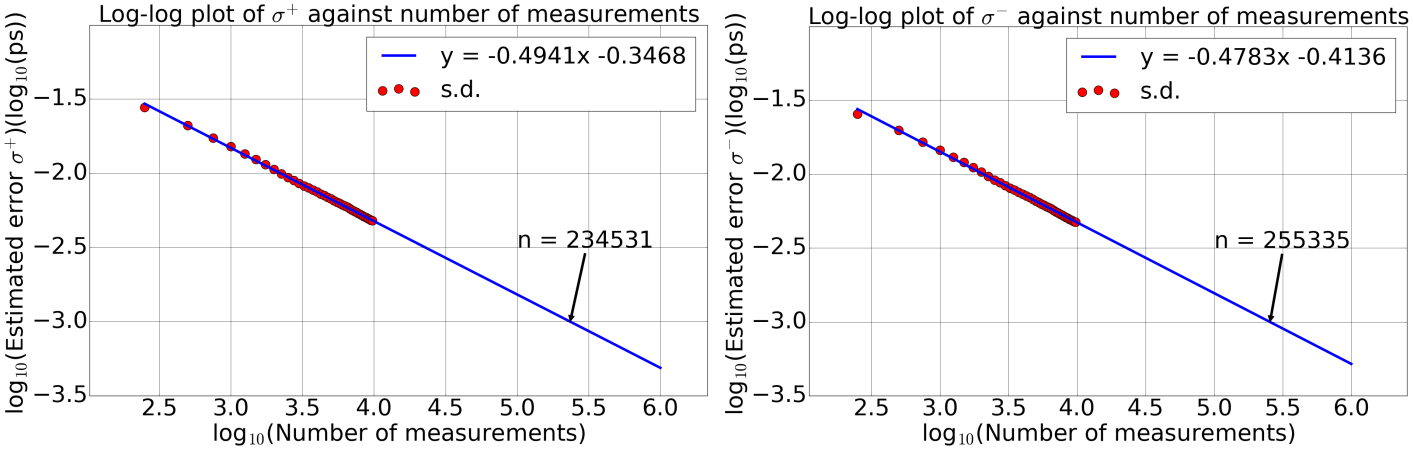

FIG. 4: The dependence of the standard deviation on the number of measurements in logarithmic scales. The minimisation of NLL function took initial guesses of 0.2ps, 0.3ps and 0.5ps. Each figure depicts a linearly decreasing pattern of the standard deviation with the number of measurements in logarithmic scales. Thus a linear fit was applied and it was extrapolated, assuming the pattern stayed linear in the region of interest. The extrapolation yielded the required number of measurements for an accuracy of 10^{-15}s as (2.3 to 2.6)$\times 10^5$.

ods. With the tolerance level set to 10^{-6}, the results suggest that the parabolic estimate and the NLL function is barely distinguishable within $\pm 1\sigma$. Nevertheless, it was found that the magnitude of σ^+ was slightly larger when using the NLL function and considering down to 5 d.p. This was expected due to the asymmetry of the function (FIG. 3). Therefore, error estimations using the NLL function are expected to be more plausible, especially when one chooses to investigate higher standard deviation. On the other hand, parabolic estimation assumes perfect symmetry, becoming increasingly invalid away from the minimum. Yet, the latter provides a simple, efficient means of estimating the uncertainty that is robust while generating relatively accurate approximations. Furthermore, the reader should be aware that the former method is also an approximation and becomes exact only in the limit of the dataset $N \to \infty$.

With the above results generated computationally, the dependence of the standard deviation on the number of measurements was investigated. The error estimation using $|\Delta\text{NLL}| = \frac{1}{2}$ was employed for this part of the investigation due to the reasons discussed above, and only the secant method was utilised to identify τ^+ and τ^- due to its speed of convergence. The initial size of the dataset was 250 and the procedure was repeated while increasing the size by 250 each time. Subsequently, the standard deviation was plotted against the number of measurements in logarithmic scales. The results are shown in FIG. 4.

Both graphs in FIG. 4 depict a linear pattern on logarithms scales, suggesting an exponential decrease on linear scales. There appears to be a general decreasing pattern of the standard deviations with the number of measurements. This is consistent with the expectation that the variance goes as $\sim \frac{1}{N}$ due to the central limit theorem. From the extrapolation of the linear fits, it was found that the number of measurements required to achieve an accuracy of 10^{-15}ps is between (2.3 to 2.6)$\times 10^5$; that is ~ 24 times the size of the data used in this analysis.

To summarise this part of the analysis, neglecting the possible background signals yielded the average lifetime of D^0 meson as (404.5$\pm$4.7)$\times 10^{-15}$s, showing a deviation of approximately 6×10^{-15}s from the value provided by the Particle Data Group[4], which is (410.1$\pm$1.5)$\times 10^{-15}$s. Hence, ignoring the possible background signals appears to have skewed the result. The following subsection will provide the details of the analysis carried out by taking the background into account.

4.2 With background signals

4.2.1 Minimisation of the 2D NLL

Prior to minimising the 2D NLL function defined in Eq. (10a), the computational methods written, namely the gradient, Newton's and the Quasi-Newton methods with DFP algorithm, were tested on a 2D hyperbolic cosine function. The results are illustrated in FIG. 5. The figures demonstrate each iteration converging towards the minimum point until the change in the parameters became smaller than the tolerance level of 10^{-6}. The estimations of the minimum position are of the order of 10^{-5} for the gradient and the Quasi-Newton while Newton's method generated smaller values of orders 10^{-13} and 10^{-8}, suggesting that all three methods are adequate in optimising the 2D NLL function.

The outcome of the minimisation of the 2D NLL is shown in FIG. 6 and the numerical results are summarised in TABLES II and III. The numerical results and the graphical outputs are consistent with each other. The values of the best fit parameters τ_{min} and a_{min} from all three methods are equal correct to 4 d.p. From the simulations, Newton's method demonstrated the fastest converging speed on both the hyperbolic cosine function and the NLL function, whereas the other two appeared to converge at a similar speed; the Quasi-Newton being slightly faster on the hyperbolic cosine function. Nevertheless, at higher dimensions the overall computational time associated with Newton's method is expected to dramatically increase as it requires computing both the first and second derivatives of the multi-dimensional function as well as the inverse of the Hessian matrix. Moreover, during the simulations, Newton's method demonstrated sensitivity on the initial parameters, returning invalid iterations unlike the other two methods. This was expected as calculating the Hessian using finite differences often yields a matrix that is not strictly positive definite, leading to (Eq. (15)) not being satisfied.

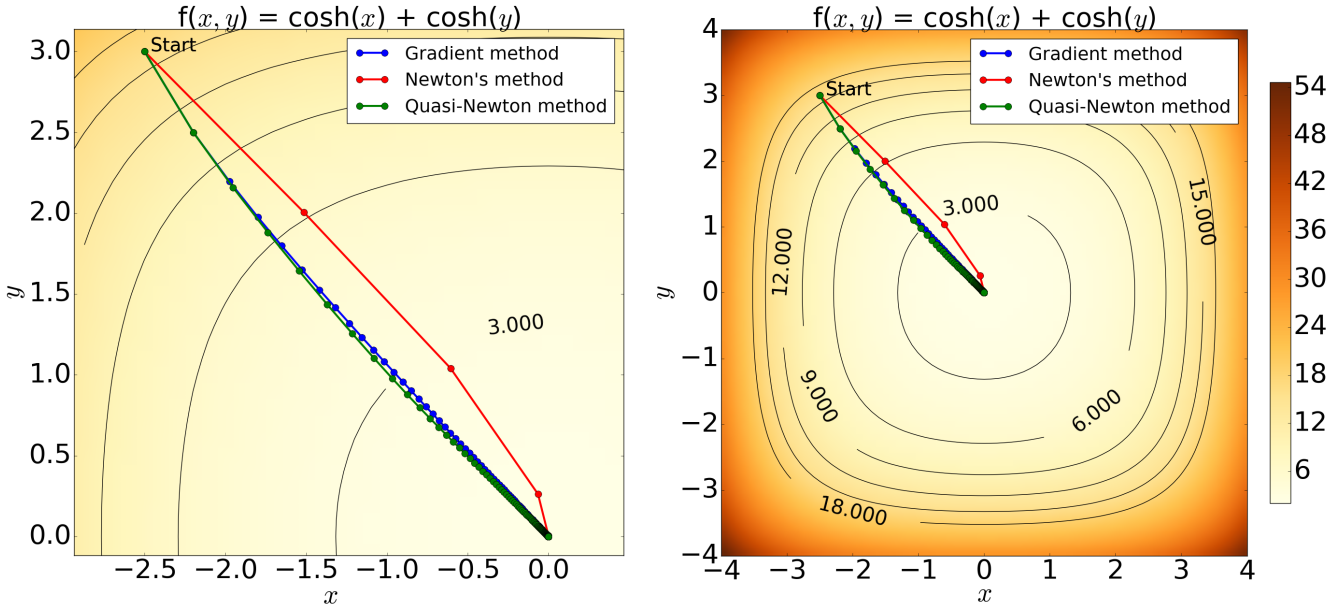

FIG. 5: Contour plots of the 2D hyperbolic cosine function showing the result from the minimisation with an initial condition of $(x,\ y) = (-2.5,\ 3.0)$, step-length of $\alpha = 0.05$ and a tolerance level of 10^{-6}. The left figure is an enlarged version of the right. The minimum estimated using the Quasi-Newton, gradient and Newton's methods are: $(x, y) = (-1.92,\ 1.91) \times 10^{-5}$, $(x, y) = (-1.86,\ 1.96) \times 10^{-5}$ and $(x, y) = (-2.42 \times 10^{-13},\ 6.72 \times 10^{-8})$ with 213, 222 and 5 iterations, respectively. The results graphically demonstrate the minimisation process with all the methods yielding expected results and thus confirming the validity of the computation. The paths generated by the Quasi-Newton and the gradient methods show only a small difference with similar number of iterations, whereas Newton's method illustrates a greater converging speed.

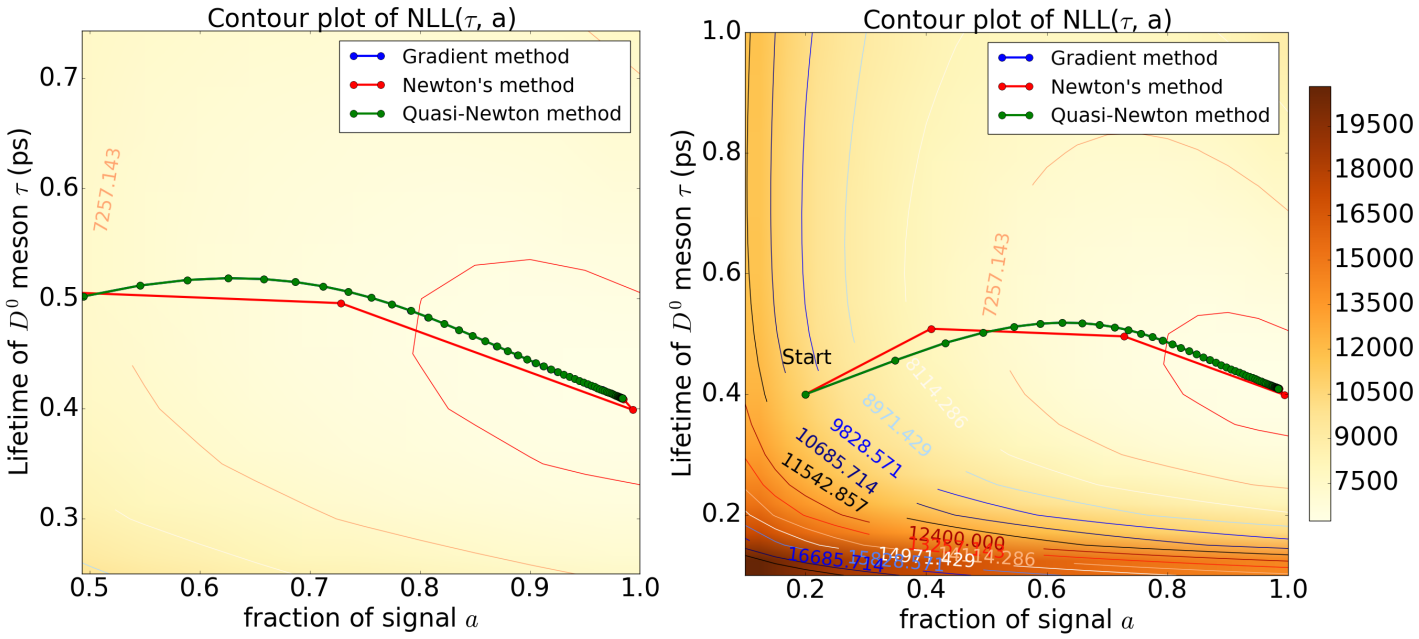

FIG. 6: Contour plots of the 2D NLL function showing the result from the minimisation with initial condition of $(a,\ \tau) = (0.2,\ 0.4\text{ps})$, step-length of $\alpha = 0.00001$ and a tolerance level of 10^{-6}. The plot of the left is an enlarged version of the plot on the right. The positions of the minimum estimated using the Quasi-Newton, gradient and Newton's methods were identical correct to 4 d.p. The estimated position of the minimum is $(a,\ \tau) = (0.9837,\ 0.4097\text{ps})$ with 98 iterations for the first two methods and 6 for the third. The figures show that the paths taken during the minimisation process are almost identical for the Quasi-Newton and the gradient method; the blue curve virtually superimposes the green curve. The path generated by Newton's method, on the other hand, differs and identifies the minimum in relatively small number of iterations. Note: CDS was used to approximate the gradients for this particular result.

TABLE II: Summary of the error results from 2D minimisation quoted correct to 4 d.p.

Method		$a_{min} = 0.9837$		$\tau_{min} = 0.4097$ps	
		σ^+ (ps)	σ^- (ps)	σ^+ (ps)	σ^- (ps)
Gradient	Bisection:	0.0074	0.0076	0.0048	0.0048
	Secant:	0.0074	0.0076	0.0048	0.0048
Newton's	Bisection:	0.0074	0.0076	0.0048	0.0048
	Secant:	0.0074	0.0076	0.0048	0.0048
Quasi-Newton	Bisection:	0.0074	0.0076	0.0048	0.0048
	Secant:	0.0074	0.0076	0.0048	0.0048

TABLE III: Summary of the results from the 2D minimisation quoted correct to 4 d.p.

Method:	Gradient	Newton's	Q-Newton
Avg. lifetime τ_{min} (ps):	0.4097	0.4097	0.4097
Fraction of signal a_{min}:	0.9837	0.9837	0.9837
Fraction of background signal:	0.0163	0.0163	0.0163
No. of iterations	98	6	98

4.2.2 Uncertainty in τ_{min} and a_{min}

The uncertainty values in TABLE II were calculated using the method stated in Section 3.3. The assumption made in this calculation was that τ and a are uncorrelated. Therefore, the identical error analysis was carried out, keeping $\tau = \tau_{min}$ constant while estimating the uncertainty in a and vice versa. However, it is evident from the contour plot (FIG. 6) that there is a degree of diagonality and thus some level of correlation must exist between the two parameters. In addition, the $\mathbf{G} \simeq \mathbf{H}^{-1}$ matrix utilised in the Quasi-Newton method showed non-zero diagonal elements, suggesting that there exist some correlation between the parameters. Therefore, the magnitude of the uncertainties in TABLE II were concluded to be underestimated.

In order to account for the possible correlation between the two parameters, Eqs. (26) and (27) were employed to compute the elements in the error matrix $\mathbf{E}$ (Eqs. (25)). This yielded $\mathbf{E}$ as:

$$\mathbf{E} = 10^{-5} \begin{pmatrix} 3.02 & -2.26 \\ -2.26 & 7.33 \end{pmatrix}, \tag{28}$$

correct to 3 s.f. The uncertainties and the correlation coefficient corresponding to the error matrix are summarised in TABLE IV.

TABLE IV: Summarising the final result with the uncertainties from the error matrix

Final results	
Avg. lifetime (ps):	0.4097 ± 0.0055
Fraction of signal:	0.9837 ± 0.0086
Fraction of background signal:	0.0163 ± 0.0086
Correlation coefficient:	-0.4813

The results are consistent with the theory. The correlation coefficient was computed as -0.4813 (4 d.p.). This result indicates that τ and a are dependent on each other. This can be observed in FIG. (7), where the ellipse is rotated with respect to the axes. It is also clear from the figure that the previous error analysis underestimated the uncertainties since $\sigma^{\pm}_{\tau,\,a} < \sigma'_{\tau,\,a}$, where $\sigma'_{\tau,\,a}$ are the errors calculated using the error matrix, representing the full range of τ and a, whereas $\sigma^{\pm}_{\tau,\,a}$ represent only a fraction of the region bounded by the error ellipse (a 1σ contour). However, there are regions outside the error ellipse where the NLL function is greater than NLL = 6219 but less than the square region spanned by the new uncertainties. Hence, the new uncertainties must be quoted with the correlation coefficient.

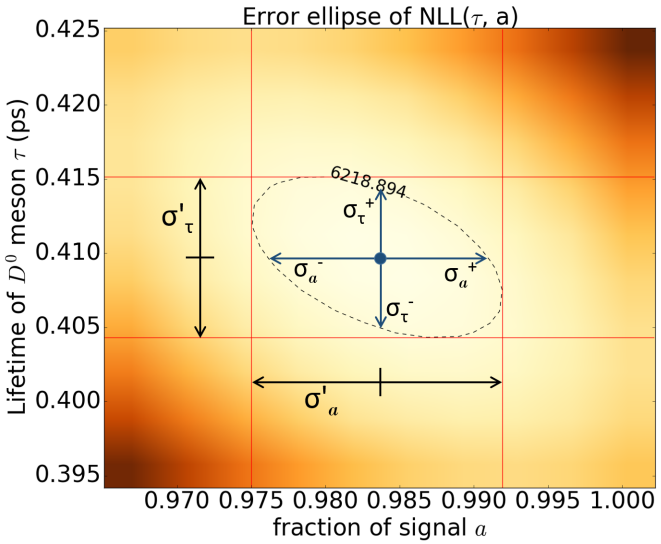

FIG. 7: The error ellipse - a contour plot corresponding to one standard deviation change in the parameters above the minimum. The notation $\sigma^{\pm}_{\tau,\,a}$ and $\sigma'_{\tau,\,a}$ are errors calculated assuming $\rho_{\tau,a} = 0$ and $\rho_{\tau,a} = -0.4813$ between τ and a, respectively, where $\rho_{\tau,a}$ is the correlation coefficient.

4.2.3 Comparing the result with the literature value

The average lifetime of D^0 meson and the fraction of the background signal were calculated to be $(409.7 \pm 5.5) \times 10^{-15}$s and 0.0163 ± 0.0086, respectively. The average lifetime is closer to the literature value provided by the Particle Data Group with a percentage difference of $\sim 0.098\%$. When considering the associated errors, the literature value lies within the calculated uncertainty, confirming the presence of the background signals which makes up around 0.0163 ± 0.0086 of the data. Therefore, by taking the limitation of the resolution of the detector into account, a more accurate result was obtained.

5. Conclusion

To summarise, a computational analysis was carried out in order to extract the average lifetime of D^0 meson in 1D and 2D cases, where the latter involved extracting the fraction of the background signal in the data in addition to the average lifetime. This was achieved by minimising the negative log-likelihood function by employing

the parabolic method for 1D, and the gradient, Quasi-Newton and Newton's methods for the 2D problem. Subsequently, the associated uncertainties were calculated by utilising the NLL function and the curvature of the last parabolic estimate, where the bisection and the secant methods were used to identify the position of τ^+ and τ^-. Furthermore, the correlation of τ and a was considered from which more realistic uncertainties and the correlation coefficient were computed.

With a tolerance level of 10^{-6}, the average lifetime was estimated to be $(404.5\pm4.7)\times10^{-15}$s in the absence of the background signals, showing a deviation of $\sim 6\times10^{-15}$s from the literature value. By accounting for the possible background signals, however, the average lifetime and the fraction of the background signals in the data was calculated to be $(409.7 \pm 5.5) \times 10^{-15}$s and 0.0163 ± 0.0086, respectively, yielding a closer value to the literature value with a percentage difference of $\sim 0.098\%$. The uncertainties quoted here was estimated using the error matrix, where the correlation coefficient was found to be -0.4813. The literature value was also found to be within the estimated uncertainties. These results verify the presence of the background signals in the data, while validating the theory of the expected distribution derived by assuming the background signals as a Gaussian due to the smearing effect caused by the limitation of the resolution of the detector.

Further investigation employing the Monte Carlo minimisation could be implemented, especially for the 2D problem, introducing some randomness in the search for the minimum. Despite the problem in hand considers only two unknown parameters, there is a large number of data associated with the analysis, subsequently leading to huge computational time that is required when searching for the optimal configuration. Therefore, Monte Carlo method could be a more efficient method in estimating the position of the minimum in the limit of the dataset $N \to \infty$.

[1] E. v. Sebille, Y. Uchida, *Project B1: A Log Likelihood fit for extracting the D^0 lifetime.*, Computational physics, 2016, Department of Physics, Imperial College London, p.1-3

[2] E. v Sebille, Y. Uchida, C. Contaldi, R. J. Kingham, *Computational Physics 2016-17 Course Note*, Department of Physics, Imperial College London, 2016, Chp.2 & 6

[3] P. Dauncey, *Second Year Statistics of Measurement Course note*, Imperial College London, Department of Physics, 2016, Chap. 3, 4, 6, 8

[4] C. Patrignani et al., *D^0 mass*, Particle Data Group, Chin. Phys. C, 40, 100001, 2016, p.1