A Log Likelihood Fit for Extracting the D^0 Meson Lifetime

December 15, 2018

Abstract

The average lifetime of the D^0 meson was determined by the minimisation of the negative log-likelihood of the decay function from the experimental data containing 10^5 measurements of the decay time and error. The behaviours of the decay functions were explored and two minimisers were developed. The result achieved with the background signals included is $0.4097 \pm 0.0048 ps$, which is in good agreement with the value given by the Particle data group.

1 Introduction

The D^0 meson is the lightest charm meson comprised of a charm quark and an anti up quark ($c\bar{u}$). There exist hundreds of different modes through which D^0 can decay via the weak interaction, mostly into some combinations of kaons and pions [1]. The average lifetime among all paths of decay is an important property of unstable particles and it can be extracted by analysing the experimental data of the decay times and the associated uncertainties.

The objective of this report is to evaluate the average lifetime of D^0 from the 10^5 measurements provided through minimising the negative log-likelihood for the decay function. A simplified case considering only the decay events is examined first as a preliminary test via the parabolic method. A more comprehensive analysis with background signals included is then completed using the Quasi-Newton method.

2 Theory

The distribution of the decay time measurements can be described by a decay function and the average lifetime can be estimated through pinpointing the minimum of its negative log-likelihood.

2.1 Decay Functions

The measured decay time t of each D^0 meson was calculated from its momentum, as well as its displacement between the points of production and decay. The uncertainties related to these measurements were propagated to give the error of decay time measurement σ. Due to the presence of this associated error, the probability density function of t is expected to be the convolution of the theoretical decay function $f^t(t)$ with a Gaussian $G(\sigma)$. The true decay function $f_{sig}^m(t)$ for the signals detected from decay events can be expressed as:

$$f_{sig}^m(t) = f^t(t) * G(\sigma) = \frac{1}{2\tau}\exp\left(\frac{\sigma^2}{2\tau^2} - \frac{t}{\tau}\right)\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau} - \frac{\tau}{\sigma}\right)\right) \tag{1}$$

Where τ is the lifetime.

Inevitably, there exists some random background signals that could not be eliminated from the measurements. The background function is the convolution of a delta function with a Gaussian and has a form of:

$$f_{bkg}^m(t) = \delta(t) * G(\sigma) = G(\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-t^2}{2\sigma^2}\right) \tag{2}$$

Considering both the decay events and the random background, the total measurement function $f^m(t)$ that most accurately represents to the distribution of t can be written as:

$$f^m(t) = a f_{sig}^m(t) + (1-a) f_{bkg}^m(t) \tag{3}$$

where a is the fraction of the decay events in the the measurements.

2.2 Negative Log Likelihood

The PDF $P(\tau; t_i, \sigma_i)$ for the D^0 meson decay experiment depends on an unknown parameters lifetime τ. With the knowledge of the decay function, τ can be determined from the experimental data of t and σ.

The likelihood function of n decay time measurements is defined as $L(\tau) = \prod_{i=1}^{n} P(\tau; t_i, \sigma_i)$, where $P(\tau; t_i, \sigma_i)$ is the decay function. The τ value at the maximum in the likelihood function has the greatest probability of obtaining the observed data and therefore is a good estimate for the true average lifetime τ.

Instead of the likelihood function itself, the negative log-likelihood function NLL was used for the investigation of τ:

$$\mathrm{NLL}(\tau) = -\log\left(\prod_{i=1}^{n} P(\tau; t_i, \sigma_i)\right) = -\sum_{i=1}^{n} \log P(\tau; t_i, \sigma_i) \tag{4}$$

Conventionally, it is preferred to use the log-likelihood function in order to decrease the value returned by the likelihood function by exploiting the property that the maximum point remains the same upon the transform. Since most optimisation algorithms are minimisers, a negative sign is added to reverse the interest and facilitate the analysis[2]. For NLL, one standard deviation is the range which shifts the value away from the minimum by +0.5.

3 Method

The true decay function $f_{sig}^m(t)$ with only one variable τ was considered first using the parabolic method. The examination of the total measurement function $f^m(t)$ with two variables τ and a was then completed by applying the Quasi-Newton algorithm.

3.1 Parabolic Minimisation

The parabolic method takes advantage of the property that most functions can be approximated as a parabola close to a minimum. The minimum point x_{min} for a function $f(x)$ can be estimated by successively fitting a 2$^{\text{nd}}$ order Lagrange polynomial $P_2(x)$ to three points (x_0, x_2, x_2) near the minimum and differentiating $P_2(x)$ to find the minimum point x_3 for this interpolation. The calculation of x_3 can be done by using the equation derived from $\frac{dP_2}{dx}|_{x=x_3} = 0$:

$$x_3 = \frac{1}{2}\frac{(x_2^2 - x_1^2)f(x_0) + (x_0^2 - x_2^2)f(x_1) + (x_1^2 - x_0^2)f(x_2)}{(x_2 - x_1)f(x_0) + (x_0 - x_2)f(x_1) + (x_1 - x_0)f(x_2)} \tag{5}$$

At the end of each iteration, the three points with the lowest values out of the known four are assigned to be the new input. The parabolic approximation improves as it approaches the minimum, therefore the estimate x_3 converges towards x_{min} in the process.
This method has a high convergence rate and does not depend on derivatives along the function, therefore it is a suitable 1D minimiser for the true decay NLL with one variable lifetime τ.
The algorithm was tested by using the hyperbolic function $f(x) = cosh(x)$. Given an input $x_0 = -1$, $x_1 = 0$, $x_2 = 1$, the output returned was $x_{min} = 0$. The efficacy of the algorithm was validated, but the degree of precision is controlled by the convergence criteria.
The implementation for the search of the average lifetime τ from the true decay distribution f_{sig}^m can be outlined as:

1. The points $\tau_0 = 0.35$, $\tau_1 = 0.4$, $\tau_2 = 0.45$ were selected as the initial input. From the NLL graph (Figure 5), it can be seen that the minimum point lies in close proximity to 0.4, hence these points were chosen to reduce the time needed for convergence.

2. The estimate minimum point τ_3 was calculated.

3. The lowest three points out of NLL(τ_0), NLL(τ_1), NLL(τ_2), NLL(τ_3) were used as the entry for the next iteration.

4. The step 2 and 3 were repeated until τ_3 converged sufficiently to conclude that the minimum point τ was found. The convergence criteria were set to end the process when the range of the three points is less than 10^{-5}. This requirement allows the result to be accurately quoted to a precision comparable to the value provided by the Particle Data Group.

3.2 Quasi-Newton Minimisation

The Newton method is an iterative algorithm for investigating the minimum point $\vec{x}_{min}$ for a multi-dimensional function $f(\vec{x})$ by utilising the gradient and the curvature at each step. Although the algorithm ensures fast convergence, the computation of the Hessian matrix $\mathbf{H}$ describing all the partial second derivatives might be difficult.
Alternatively, the Quasi-Newton method applies the approximation to the inverse Hessian $\mathbf{G}_n$ for including the curvature in the calculation. This approach is more efficient as it does not involve inverting $\mathbf{H}$ through solving matrix equation. The minimisation can be achieved by consecutively uating the estimate

$$\vec{x}_{n+1} = \vec{x}_n - \alpha \mathbf{G}_n \cdot \vec{\nabla} f(\vec{x}_n) \tag{6}$$

where α is a parameter much less than 1.
$\mathbf{G}_0$ is an $N \times N$ identity matrix depending on the number of variables N subjected to minimisation.

$\mathbf{G}_n$ at each iteration can be updated by applying the Davidon-Fletcher-Power algorithm

$$\mathbf{G}_{n+1} = \mathbf{G}_n + \frac{(\vec{\delta}_n \times \vec{\delta}_n)}{\vec{\gamma}_n \cdot \vec{\delta}_n} - \frac{\mathbf{G}_n \cdot (\vec{\delta}_n \times \vec{\delta}_n) \cdot \mathbf{G}_n}{\vec{\gamma}_n \cdot \mathbf{G}_n \cdot \vec{\gamma}_n} \tag{7}$$

where $\vec{\delta}_n = \vec{x}_{n+1} - \vec{x}_n$ and $\vec{\gamma}_n = \vec{\nabla} f(\vec{x}_{n+1}) - \vec{\nabla} f(\vec{x}_n)$.
The effectiveness of the algorithm was verified with the function $f(x,y) = (x-1)^2 + (y-3)^2$. For $(x_0, y_0) = (2,2)$ and $\alpha = 5 \times 10^{-2}$, the result produced was $(x_{min}, y_{min}) = (1,3)$, with a precision depending on the convergence criteria. As a result, the estimate generated by the algorithm can be expected to be in good agreement with the true value.
The procedure of acquiring the average lifetime τ and the fraction a from the total measurement function f^m can be summarised as:

1. Several choices for initiating the process were made. $(\tau_0, a_0) = (0.4, 1.0)$ was used as the starting point because of its closeness to the minimum as seen in the 3D NLL graph (Figure 8). α was set to be 10^5 to give high operation speed and accurate result.

2. The new estimates τ_n and a_n were calculated.

3. The new gradient $\vec{\nabla} f^m(\tau_n, a_n)$ was derived from the first principle

4. The new inverse Hessian approximation G_n was evaluated.

5. The step 2 to 4 were repeated until τ_n and a_n both converged sufficiently to conclude that the minimum point (τ, a) was found. The convergence criteria were used to end the process when the difference between two successive steps is less than 10^{-6} for both quantities so that the result can be quoted to the same precision as the data retrieved from PDG.

4 Errors, Results and Discussion

4.1 Experimental Data

The data contains 10^5 measurements of the decay time and the associated error (t, σ) for the D^0 meson in the unit of picosecond ps.
The probability density distributions of t and σ are displayed in Figure 1. It was found that 99.9% of the decay time measurements t lies between -1.5 and 3ps with negligible outliers, and the distribution is slightly skewed to the right. When compared to the fitted function $f^m_{sig}(t;, \sigma = \sigma_{avg}, \tau = t_{avg})$, the similarities in the overall shapes and positions suggest that the distribution of t conforms to the true decay function $f^m_{sig}(t)$, and $t_{avg} = 0.4034ps$ can be considered as a reasonable first guess for the average lifetime τ.
On the other hand, the measurement errors σ has a skewed right, truncated distribution with a spread between 0 and 0.6ps. The boundary at 0 is sensible because σ was defined as the standard deviation of a Gaussian about the corresponding t, therefore it must be positive. The abrupt drop on the right can potentially be explained by some experimental or statistical measures taken before the analysis to keep the uncertainties below a specified threshold.

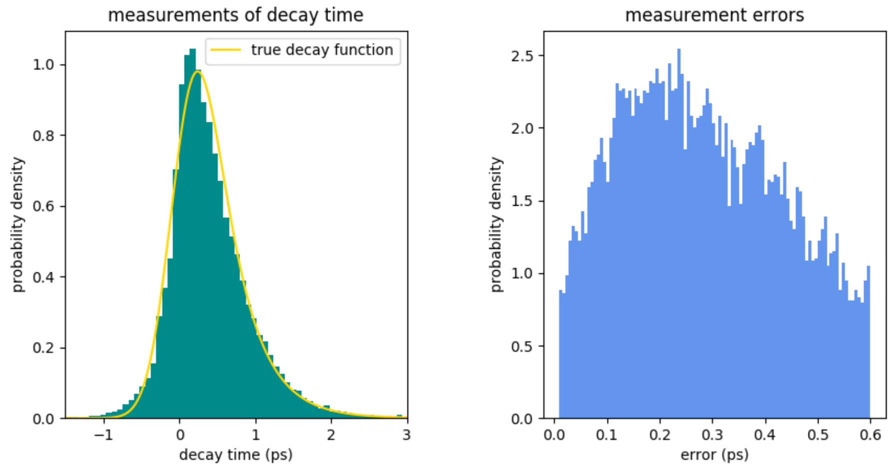

Figure 1: Histograms of the data of decay time and error

4.2 Average Lifetime from the True Decay Function

Assuming that the background signals are negligible compared to the detected decay signals, the average lifetime τ extracted from the true decay function $f^m_{sig}(t)$ is expected to be a good estimate. Some preliminary analysis was conducted prior to the minimisation in order to ensure the behaviours and properties of $f^m_{sig}(t)$ are thoroughly understood.

In Figure 2, the behaviours of $f^m_{sig}(t)$ upon changing σ and τ were examined respectively. It was observed that increasing σ and τ both results in the flattening and broadening of the curve, as well as the shift of the peak towards the right.

For greater σ, the Gaussian width related to t measurement is larger. The convoluted $f^m_{sig}(t)$ thus has a peak moved further away positively from 0 and the range of t is consequently widen on both side of the centre. In contrast, an increase in τ only brings the peak to and stretches the tail on the right towards larger t. This can be justified as a longer lifetime τ naturally leads to more possible and greater values of decay time t.

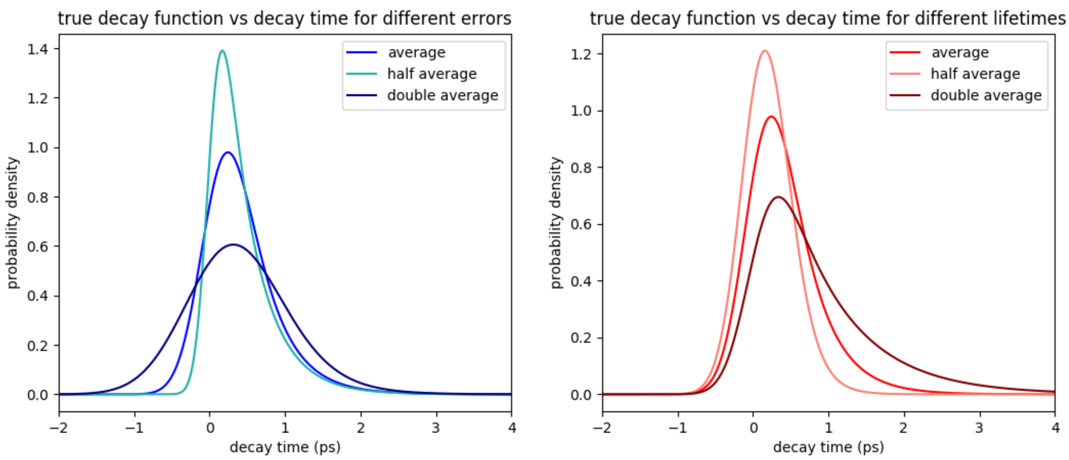

Figure 2: The true decay function against decay time for different errors

The integral over t should always be unity because $f^m_{sig}(t)$ is a PDF. An integration algorithm using the extended trapezoidal rule was implemented to verify this property and the integrals across different σ and τ are shown in Figure 3. The result confirms that the integral with respect to t is invariant with a value of 1 and is independent of σ and τ.

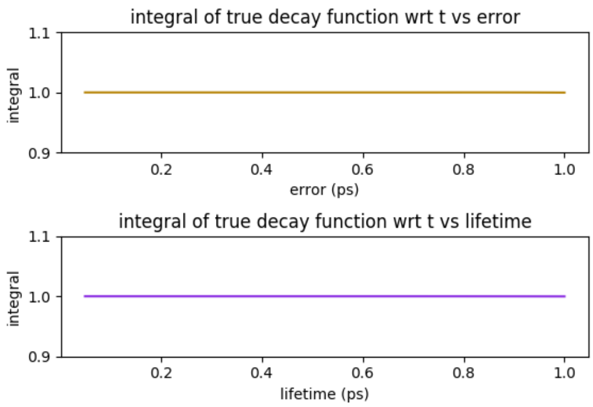

Figure 3: Integrals of the true decay function over decay time against error and lifetime graphs

The negative log-likelihood of $f^m_{sig}(t)$ as a function of τ was plotted in Figure 6. The minimum point approximately locates at $\tau = 0.4ps$, which agrees with the initial guess. Hence the points $\tau_0 = 0.35ps$, $\tau_1 = 0.4ps$, $\tau_2 = 0.45ps$ were applied as the initial input for the parabolic minimisation algorithm (as discussed in Section 3.1). The average lifetime τ obtained is $0.4045ps$, where the NLL has the minimum value of 6220.45.

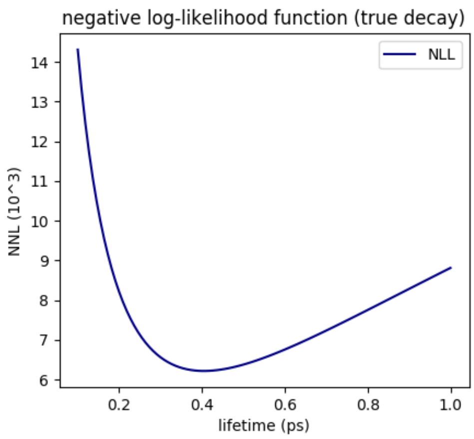

Figure 4: Negative log-likelihood of the true decay function

The standard deviation σ_τ of τ was computed through three different methods:

1. Searching for the values τ_+, τ_- where the NLL is changed by 0.5: approaching τ_+, τ_- from either side of the minimum by successively identifying the interval they position at and using progressively small intervals.

2. Evaluating the values τ_+, τ_- where the NLL is changed by 0.5 through deducing and applying the quadratic equation near the minimum.

3. Equating the Gaussian approximation of NLL to the interpolated quadratic polynomial for calculating the Gaussian width.

	method 1	method 2	method 3
σ_τ (ps)	$+0.00474, -0.00468$	± 0.00470	± 0.00471

Table 1: Standard deviations of lifetime from the true decay function

The values of σ_τ found from different approaches are presented in Table 1.
The uncertainty produced via the first method should be the most accurate among the three, because it does not rely on the parabolic approximation and it preserves the asymmetric nature of the NLL. However, it is time consuming since many values have to be trialed before reaching the end result. The other two methods both utilise the parabolic estimate, therefore the errors given are very similar. Although the uncertainty might be less accurate due to the approximation used, the operation times for these routines are significantly shorter. Considering that the results should only be recorded to $10^{-4}ps$, the latter two are preferred as they offer better efficiency for the same outcome in this case. The standard deviation σ_{tau} for τ is $\pm 0.0047ps$.
In Figure 7, the relationship between σ_{tau} and the number of measurements N included in the analysis was investigated. The values σ_τ can be reduced by increasing the data sample size. By fitting the function $\sigma_\tau \times \sqrt{N} = 0.4662$, which was derived empirically from the data, it was discovered that σ_τ is inversely proportional to $\sqrt{N}$. In order to suppress σ_τ down to below $10^{-3}ps$ $(1fs)$, it is necessary to collect at least about 217400 measurements of (t, σ) from the D^0 decay experiment.

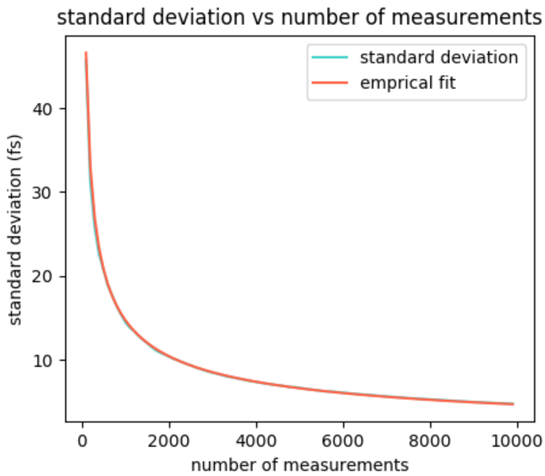

Figure 5: Standard deviation against number of measurements graph

The average lifetime τ estimated from the true decay function $f^m_{sig}(t)$ is $0.4045 \pm 0.0047ps$. This result was compared with the current most accurate measurement in the world $\tau_{PDG} = 0.4101 \pm 0.0015ps$[3] provided by PDG. Whilst the percentage difference of 1.354% is acceptable, τ and τ_{PDG} do not lie within the error range $\pm\sigma_\tau$ of each other. The acquired value τ has insufficient accuracy. It can be explained by the disregard for the background $f^m_{bkg}(t)$ and the assumption made for the application of $f^m_{sig}(t)$ for minimisation is invalid.

4.3 Average Lifetime from the Total Measurement Function

The minimisation was performed on the total measurement function $f^m(t)$, which takes the background signals into account. The negative log-likelihood of $f^m(t)$ as a function of τ and a was plotted as a 3D surface in Figure 7.

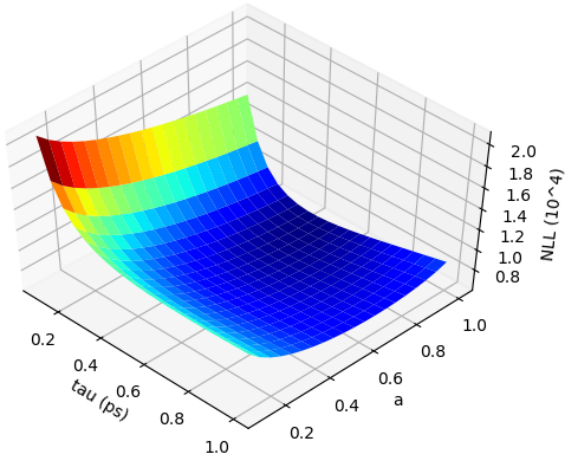

Figure 6: Negative log-likelihood of the total measurement function

From the graph, the starting point for the process was chosen to be $(\tau_0, a_0) = (0.4ps, 1.0)$ as discussed in Section 3.2. The estimates for τ and the fraction of the decay signals a were obtained through the Quasi-Newton minimisation algorithm. The minimum point was found at $(\tau_0, a_0) = (0.4097ps, 0.9837)$, where the NLL has it lowest value 6218.39.
The standard deviations σ_τ and σ_a were computed by using the methods introduced previously on two quantities individually. When the uncertainty for one quantity was being examined, it was assumed that the other is fixed to the minimum point value. The results are shown in Table 2.

	method 1	method 2	method 3
σ_τ(ps)	$+0.00485, -0.00478$	± 0.00485	± 0.00481
σ_a	$+0.00735, -0.00764$	± 0.00749	± 0.00751

Table 2: Standard deviations of lifetime and fraction from the total measurement function

For σ_τ, the results are consistent across different approaches, hence the similar argument used for the $f^m_{sig}(t)$ case was applied to determine that $\sigma_\tau = \pm 0.0048ps$. Conversely, the difference between the values of σ_a from the first method suggests that the parabolic approximation does not hold very well for $f^m(t)$ with respect to a. The mean value of σ_{a_+} and σ_{a_-} for the first method was taken, and it turned out to be compatible with the results from the other two routines, thus $\sigma_a = \pm 0.0075$.
The average lifetime τ evaluated from the total measurement decay function $f^m_{sig}(t)$ is $0.4097 \pm 0.0048ps$. This value gives a small percentage difference of 0.975% relative to τ_{PDG}, and it lies within the uncertainty range of τ_{PDG}. It can be concluded that the estimate of τ achieved is in good agreement with the PDG data.
The fraction of the signals due to the decay events a is 0.9837 ± 0.0075, meaning that the background only contributes to 1.630% of the total signal detected. The values of τ with and without incorporating the background signals only differ slightly by 1.253%. Even though the change is not significant, it is still essential for attaining better accuracy.
In summary, the average lifetime τ extracted from the experimental data of the D^0 meson decay can be presented in Table 3.

fit function	average lifetime τ (ps)
true decay function $f^m_{sig}(t)$	0.4045 ± 0.0047
total measurement function $f^m(t)$	0.4097 ± 0.0048

Table 3: Average lifetime of the D^0 meson

5 Conclusion

The average lifetime of the D^0 meson acquired from the true decay function is $0.4045 \pm 0.0047 ps$, which does not overlap with the error range for the current best value quoted from the Particle Data Group due to the negligence of the background signals. By including the background in the total measurement function, it was discovered that the background counts towards 1.630% of the observed signals, and the average lifetime obtained has a value of $0.4097 \pm 0.0048 ps$. The accuracy of this result was confirmed by the good agreement with the PDG measurement.
Further improvements can be accomplished by collecting more measurement data to reduce the standard deviation, as well as modifying the parameters and convergence criteria used in the minimisation algorithms to reach higher precision.

References

[1] Abe, K. et al. (1984). Measurement of the Lifetime Difference in D0 Meson Decays. *Physical Review Letters*, 88(16).

[2] Press, W. (1986). *Numerical Recipes*. 3rd ed. Cambridge: Cambridge University Press.

[3] Particle Data Group (2018). *CHARMED MESONS*. [online] Particle Data Group. Available at: http://www-pdg.lbl.gov/2015/tables/rpp2015-tab-mesons-charm.pdf [Accessed 15 Dec. 2018].