Computational Physics: Project B1

Christopher Olley, Imperial College London

December 22, 2017

Word Count: 2181

Abstract

This project investigates data from the decay of D_0 mesons in order to produce a lifetime estimate $\hat{\tau}$ using the negative log-likelihood (NLL) method. Initially a 1 dimensional model is used, approximating the distribution of measured lifetimes by a exponential function convoluted with a Gaussian. A 2 dimensional distribution is then used, introducing a parameter a that takes into account the proportion of background signal. Parabolic minimisation, the gradient method and Newton's method are used in order to minimise the resulting NLL functions, and the resulting estimates are compared with data from Olive et. al. (2014). The 1 dimensional model results in an estimate of $\hat{\tau} = (404.546^{+4.738}_{-4.675}) \times 10^{-15}$ s, corresponding to a percentage uncertainty of 1.35%, and the 2 dimensional model yields estimates $\hat{\tau} = (409.684^{+5.549}_{-5.435}) \times 10^{-15}$ s and $\hat{a} = (0.984^{+0.008}_{-0.009}) \times 10^{-15}$ s, a percentage uncertainty of 0.101%. In the 1 dimensional case, the variation of the measurement error with the size of the data set is also examined using a linear regression, with the resulting trend obeying the standard result that the variance of the mean scales as $\sigma_{\overline{x}} \propto 1/\sqrt{N}$.

1 Investigation Aims

1.1 Introduction

The lifetime of subatomic particles is typically measured using their momentum and position, as the times involved are too short to measure directly (Griffiths 2008). Measurements of momentum and position can be related to lifetime in the following way:

$$t = m\frac{\vec{x} \cdot \vec{p}}{p^2}. \tag{1}$$

In this project, data obtained from the decay of D^0 mesons is analysed in order to produce a lifetime estimate using negative log-likelihood estimation. The data set contains lifetime measurements from 10000 decay events and an associated uncertainty for each data point.

1.2 Background

A large number of measurements of identical decay events will follow an exponential distribution in the idealised theoretical case, given by

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

where t is the measured decay time and τ is the average lifetime (Barlow 1989). The uncertainty on each measurement can be factored in through a convolution of the distribution with a Gaussian of width σ, resulting in a first approximation to the measured distribution,

$$f^m_{sig}(\tau;t,\sigma) = \int_{-\infty}^{\infty} \frac{1}{\tau}\exp\left(\frac{-t'}{\tau}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{(t-t')^2}{\sigma^2}\right)dt' \tag{3}$$

$$= \frac{1}{2\tau}\exp\left(\frac{\sigma^2}{2\tau^2}-\frac{t}{\tau}\right)\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(\frac{\sigma}{\tau}-\frac{t}{\sigma}\right)\right). \tag{4}$$

The data set in fact also contains some background signal which corresponds to a random signal smeared by the limited resolution of the detector, resulting in a Gaussian distribution shape with a mean lifetime of zero,

$$f^m_{bkg}(t,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{5}$$

A more precise, 2 dimensional version of the distribution of particle lifetime measurements is derived by introducing a new parameter a that controls the proportion of background signal, in the following way:

$$f(\tau,a;t,\sigma) = af^m_{sig}(\tau;t,\sigma) + (1-a)f^m_{bkg}(t,\sigma). \tag{6}$$

2 Methods

2.1 Negative Log Likelihood

In order to fit the data to the unknown parameter τ, negative log-likelihood estimation is used. The likelihood is defined as an ensemble of probability density functions (PDFs) $\mathcal{P}(\vec{u};\vec{m}_i)$,

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

where $\vec{u}$ corresponds to a vector of unknown parameters and $\vec{m}_i$ is a set of data points. The negative log-likelihood (NLL) is given by

$$\mathrm{NLL}(\vec{u}) = -\log(\mathcal{L}) = -\sum_{i=1}^{n}\log(\mathcal{P}(\vec{u};\vec{m}_i)). \tag{8}$$

NLL estimation consists of minimising equation (8) in order to obtain an estimate for the unknown parameters $\vec{u}$ as the 'most likely' parameters given this data set. In this investigation, $\vec{m}_i$ corresponds to the lifetime and uncertainty measurements (t_i,σ_i). In the 1 dimensional model $\vec{u}$ corresponds to the true lifetime τ. In the 2 dimensional case, $\vec{u}$ corresponds to both τ and the proportion of true signal a.

The error on this estimate, $\sigma^{\pm}_{\hat{\tau}}$ can be calculated using two methods, both covered in Barlow (1989). The NLL variation method uses the fact that the standard deviation for a NLL estimate

occurs at

$$\mathrm{NLL}\left(\hat{\tau} \pm \sigma^{\pm}\right) = \mathrm{NLL}(\hat{\tau}) + \frac{1}{2}. \tag{9}$$

The upper and lower standard deviations can differ in this method if the NLL is not exactly parabolic. The curvature method relates the standard deviation to the curvature of the NLL at $\hat{\tau}$ to $\mathcal{I}$, the Fisher information, and uses the lower Cramer-Ráo bound for variance, $V(\hat{\tau}) = \sigma^2_{\hat{\tau}} = \mathcal{I}^{-1}$, where

$$\mathcal{I} = \left(\frac{\partial^2}{\partial \tau^2} \mathrm{NLL} \right) \Bigg|_{\hat{\tau}}. \tag{10}$$

The error on a 2 dimensional NLL estimate can be derived in a similar way, by finding the contour line that satisfies the following equation $\mathrm{NLL}(\tau_{cont}, a_{cont}) = \mathrm{NLL}(\hat{\tau}, \hat{a}) + 1/2$. The errors will then given by $\sigma^+_\tau = \max(\tau_{cont})$, $\sigma^-_\tau = \min(\tau_{cont})$, $\sigma^+_a = \max(a_{cont})$, $\sigma^-_a = \min(a_{cont})$.

2.2 Finite Difference Methods

Numerical minimisation often involves calculating the gradient of a function or higher order derivatives. This is usually carried out numerically when an analytic expression is not feasible, using finite difference methods. The Central Difference Scheme (CDS) can be derived in the 1 dimensional case for a function $y(x)$ by expanding $y(x+h)$ and $y(x-h)$ about x, resulting in an estimate for the derivative $\tilde{y}'_c$,

$$\tilde{y}'_c = \frac{y(x+h) - y(x-h)}{2h}. \tag{11}$$

The expansion reveals that this estimation is accurate up to a $\mathcal{O}(h^2)$ error. Higher order derivatives can be calculated using similar methods, and the expansion to higher dimensions is straightforward using higher dimensional Taylor series expansions.

2.3 Parabolic Minimisation

In order to fit the data to the 1 dimensional distribution f^m_{sig} using NLL estimation, equation (8) is minimised with respect to τ to determine $\hat{\tau}$. The minimum is calculated numerically using parabolic minimisation (Press 2002). In this method, three starting points $\{(x_0, y_0), (x_1, y_1), (x_2, y_2)\}$ are used to define a quadratic Lagrange polynomial with a minimum x_3 given by

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

The point x_n corresponding to the largest function value y_n is then discarded, and the process of calculating the minimum of the parabola of the remaining three points is repeated until convergence. This method will converge only in a region of positive curvature.

2.4 Gradient Method

To fit the data to the 2 dimensional distribution given by equation (6), equation (8) must be minimised to estimate $\hat{\tau}$ and $\hat{a}$. This requires a 2 dimensional minimisation algorithm, such as the gradient method (Press 2002). This algorithm involves following the line of steepest descent towards

the minimum, through the following iterative procedure

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

where $\vec{x}_n = (x, y)$ is a point in the 2 dimensional parameter space and α is the step size. The gradient is evaluated using a finite difference method such as the central difference scheme, as defined in section 2.2.

2.5 Newton's Method

Newton's method is another, more efficient 2 dimensional minimisation algorithm that uses the curvature at each point in the iteration. This algorithm approximates the function by a parabolic function at each point in the iteration and finds the minimum of this parabola (Press 2002). If the function is well approximated by a parabola about the minimum the algorithm will converge to the minimum quickly. The procedure is

$$\vec{x}_{n+1} = \vec{x}_n - [\mathbf{H}(\vec{x}_n]^{-1} \cdot \nabla f(\vec{x}_n), \tag{14}$$

where $\mathbf{H}$ is the Hessian, defined as

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \tag{15}$$

for a two dimensional function $f(x, y)$. The inverse of this matrix can quickly be obtained using LU decomposition.

2.6 Convergence

In all minimisation methods discussed in this report, the following convergence criteria is used. The algorithm will run until the following inequality no longer holds:

$$\frac{f(\vec{x}_{n+1}) - f(\vec{x}_n)}{f(\vec{x}_n)} > \epsilon. \tag{16}$$

A value of $\epsilon = 10^{-6}$ is used consistently throughout this report.

2.7 Newton-Raphson Method

The Newton-Raphson method is a root-finding algorithm, used here to calculate the solutions to equation (9) in order to calculate the error on the NLL estimate using the variation method (Weisstein, E. W. 2017). The basic algorithm step is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \tag{17}$$

converging at the solution to $f(x) = 0$. A limitation of this algorithm is that it will struggle to converge for functions with repeated roots as the gradient at this root will be close to zero.

3 Results

3.1 Initial Data Analysis

Visualising the data in histogram form in figure 1 allows basic properties to be determined. The number of bins is chosen to be $\sqrt{N}$ or the nearest integer, where N is the number of data points, according to convention (Barlow 1989). The mean lifetime is calculated to be $\overline{t} = 403 \times 10^{-15}$ s and the mean uncertainty $\overline{\sigma} = 282 \times 10^{-15}$ s. These two values can be used as initial estimates for $\hat{\tau}$ and $\hat{\sigma}$ for initial analysis of the 1 dimensional distribution given by equation (3). Plotting the normalised histogram overlaid with the distribution given by (3) for different values of σ and τ close to the mean value estimates provides an initial understanding of the effect of these parameters. The distribution is a PDF, and so the integral over the independent variable t should be unity, regardless of the (τ, σ) parameters used. This is confirmed in the code for all of the parameter combinations plotted in figure 2. Figure 2a shows that σ affects the width of f_{sig}^{m}, and varying τ shifts the peak of f_{sig}^{m}, as one would intuitively expect.

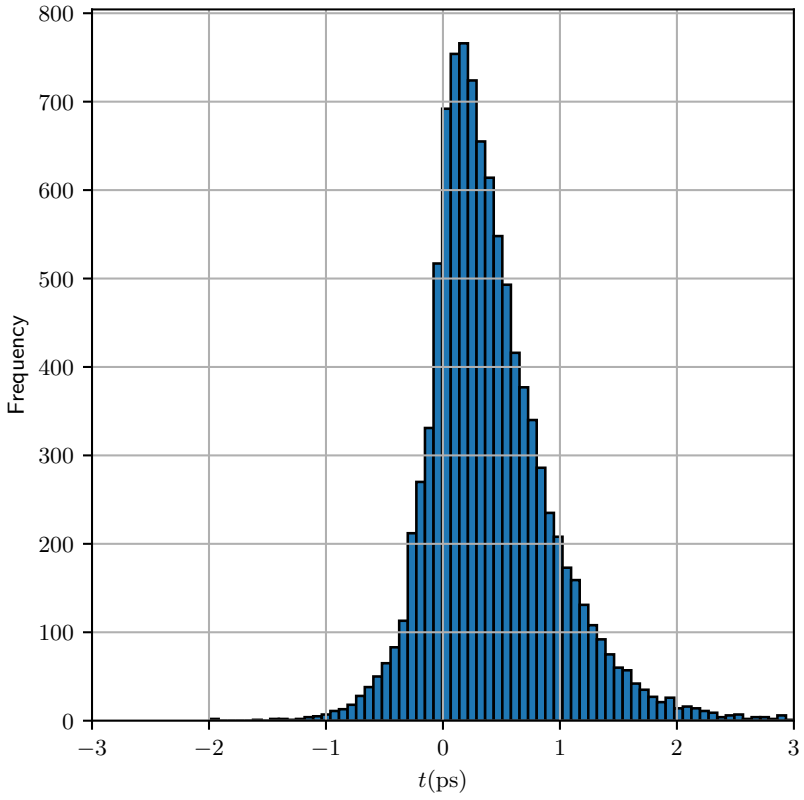
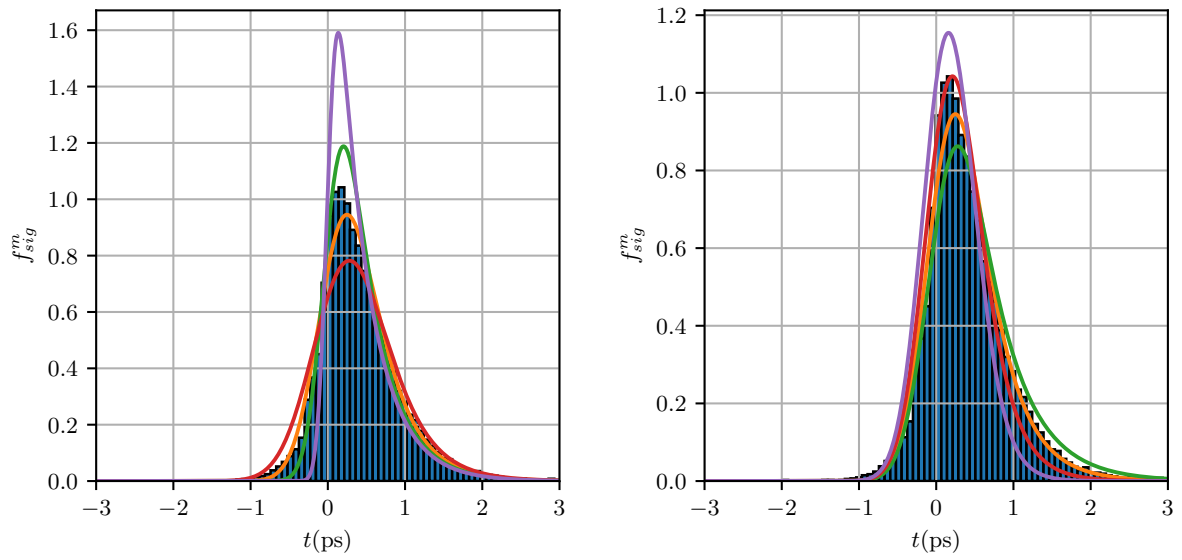


Figure 1: Histogram of particle data, plotting lifetime measurement t against frequency of measurment, using a bin number $n = \sqrt{N}$ where N is the number of data points.



(a) Variation of f^m_{sig} with σ.

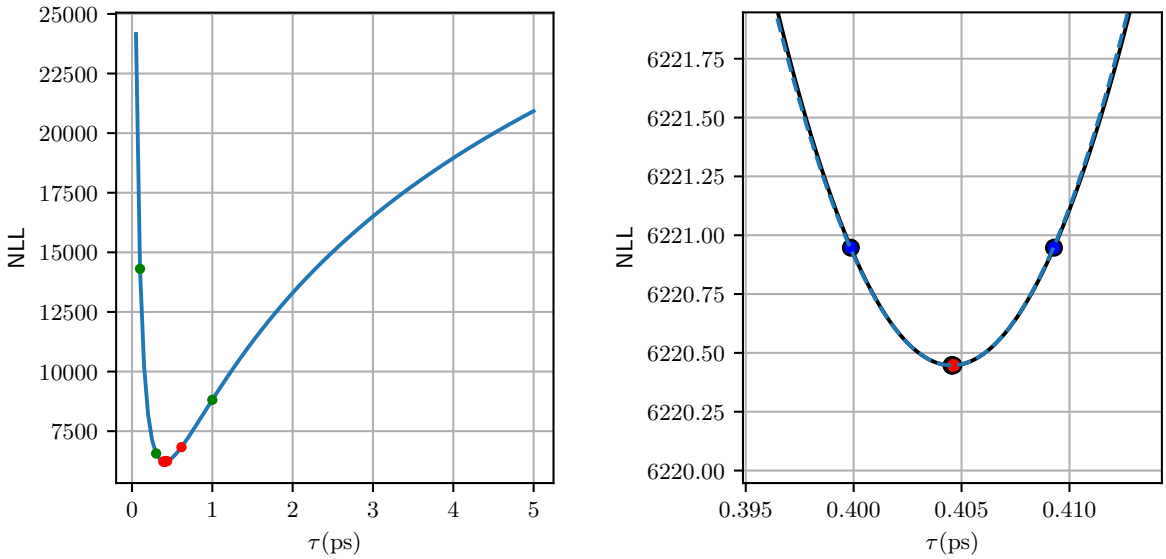
(b) Variation of f^m_{sig} with τ.

Figure 2: Comparing the variation of f^m_{sig} with τ and σ around their mean values, to the normalised histogram of decay data. In (a), $\tau = 0.4$ and $\sigma = 0.1$ (purple), $\sigma = 0.2$ (green), $\sigma = 0.3$ (orange), $\sigma = 0.4$ (red). In (b), $\sigma = 0.3$ and $\tau = 0.2$ (purple), $\tau = 0.3$ (red), $\tau = 0.4$ (orange), $\tau = 0.5$ (green) (all in ps).

3.2 One Dimensional Minimisation

The parabolic minimisation algorithm described in section 2.3 is used to minimise equation (8) using equation (3) as a distribution to determine the first NLL estimate $\hat{\tau}$. As a validation, the algorithm is used to minimise $f_{test}(x) = \cosh x$ from a variety of starting points, correctly finding the minimum $f_{test} = 1$ at $x = 0$. The parabolic minimisation algorithm applied to (8) is visualised in figure 3a for starting points $\tau = \{0.1, 0.3, 1.0\}$ps (green dots), converging at $\hat{\tau}$. The subsequent iterations of the algorithm are shown as red dots. The choice of starting points is important, as they must bracket the minimum and all be in a region of positive curvature.

The error on this estimate is calculated using the NLL variation method and the curvature method, both visualised in figure 3b. The $\sigma^{\pm}_{\hat{\tau}}$ derived using the variation method are shown as blue dots. These are calculated as the solution to the equation $\mathrm{NLL}(\hat{\tau} \pm \sigma^{\pm}_{\hat{\tau}}) = \mathrm{NLL}(\hat{\tau}) + 1/2$, using the Newton-Raphson Method, as described in section 2.1 and 2.7. The curvature used for the second method is derived from the final parabolic approximation to the NLL at the minimum (blue dotted line), which uses the final points from the parabolic minimisation algorithm (red dots). As a validation, it can be seen that the parabola closely approximates the true function at the minimum. Using the NLL variation method, the final estimate is $\hat{\tau} = (404.546^{+4.738}_{-4.675}) \times 10^{-15}$ s. Using the curvature method, an estimate of $\hat{\tau} = (404.546 \pm 4.707) \times 10^{-15}$ s is obtained. The NLL variation method is more accurate as the curvature method uses an approximation to the NLL function whereas the NLL variation method uses the true function. This can be compared to the Particle Data Group (PDG) estimate of $(410.1 \pm 1.5) \times 10^{-15}$ s (Olive et. al. 2014), resulting in a percentage difference of 1.35%, and it is seen that the two values do overlap within their uncertainties.



(a) Parabolic minimisation of the NLL.

(b) Calculating standard deviation on NLL estimate.

Figure 3: Parabolic minimisation of the NLL. In (a), equation (8) is minimised using starting points $\tau = \{0.1, 0.3, 1.0\}$ (green dots). The subsequent iterations of the algorithm are the red dots. In (b), the standard deviation $\sigma_{\hat{\tau}}^{\pm}$ is calculated by: (i) varying the value of NLL by $1/2$ (blue dots), (ii) using the curvature method based on the final parabolic approximation at the minimum (blue dotted line) using the final points from the parabolic minimisation algorithm (red dots).

3.3 Error variation with size of data set

The error on the NLL estimate, $\sigma_{\hat{\tau}}^{\pm}$, varies with the number of data points used to construct the estimate. Figure 4 visualises this dependence, plotting both the NLL variation method errors (upper and lower errors in red) and curvature method errors (green) against the number of data points N, using a log-log scale. An evenly spaced range of N values between $N = 200$ and $N = 10000$ is used. A linear regression is carried on each of the error trends, which results in a mean intercept of -0.856325 and a mean gradient of -0.488481 ± 0.001280, with correlation coefficient $r = 0.999664$. Assuming a relationship of the form $\sigma = AN^{\alpha}$, the intercept and gradient of this plot can be related to the constants A and α, resulting in $A = 0.424720 \times 10^{-12}$ s and $\alpha = -0.488481 \pm 0.001280$, i.e. a relationship between error and data set size of $\sigma = 0.424720/N^{0.488481}$ps. This is close to the relationship that would be expected from the standard result that the variance of the mean scales as $\sigma_{\overline{x}} = \sigma/\sqrt{N}$ for independent random variables with the same uncertainty σ (Barlow 1989). The value of $A = 0.424\,720$ ps therefore represents the value of the uncertainty if this was identical for each measurement. This is clearly not the case for this experiment and this value can be compared to the average of $\overline{\sigma} = 0.282$ ps calculated in section 3.1. This relationship shows that, in order to obtain an accuracy of 10^{-15} s, 225720 measurements are required, larger than the size of this data set. However an accuracy of order 10^{-15} is obtained with only 2055 measurements which is significantly smaller that the data set.

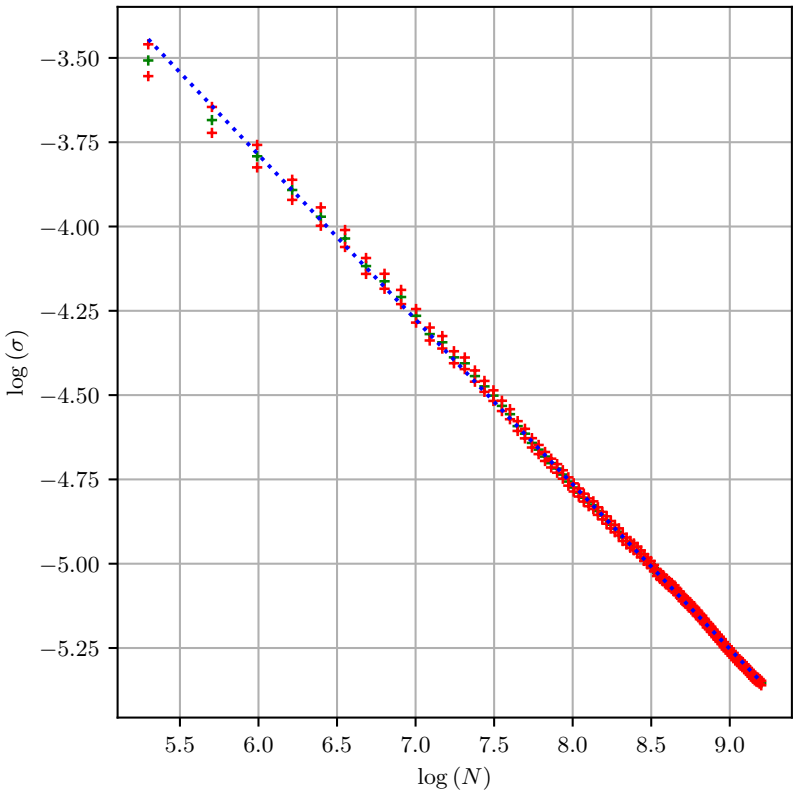
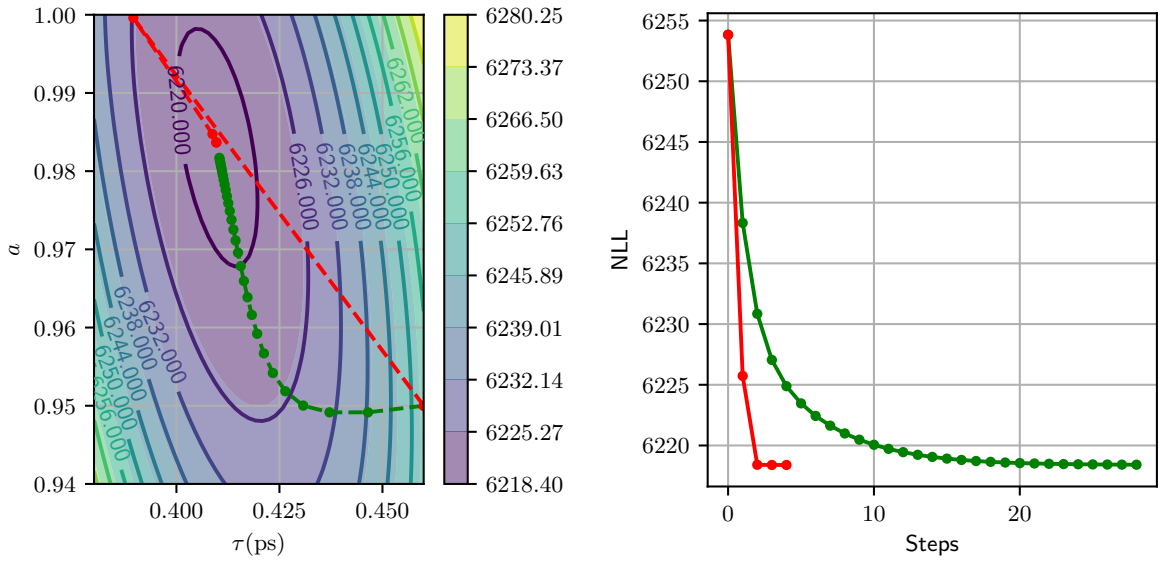


Figure 4: Log-log plot of the error on $\hat{\tau}$ given by the NLL variation method (red crosses) and the curvature method (green crosses), superimposed with the average linear regression from the three error trends. The resulting linear regression has a gradient -0.488481 ± 0.001280 and intercept -0.856325, with correlation coefficient $r = 0.999664$, implying a relationship of the form $\sigma = 0.424720/N^{0.488481}$.

3.4 Two Dimensional Minimisation

In this section the 2 dimensional minimisation algorithms described in sections 2.4 and 2.5 are applied to equation (8) using a two dimensional distribution given by equation (6). The two algorithms are validated, this time using a 2 dimensional test function $f_{\text{test}}(x, y) = \cosh(x) + \cosh(y)$, yielding the correct minimum $f_{\text{test}} = 1$ at $(0, 0)$ from various starting points. Figure 5a shows the gradient method and Newton's method applied to this minimisation problem, from a starting point $(\tau_{\text{start}}, a_{\text{start}}) = (0.46, 0.95)$ obtained from inspection of the function around the minimum. In order to ensure the gradient method doesn't 'overshoot' it is important to pick a good step size α, chosen to be 10^{-5} in this case as the largest value that doesn't result in overshooting. However this also means that the gradient method requires more iterations to converge, in this case 29 iterations for the gradient method vs 5 iterations for Newton's method, as shown in figure 5b. The gradient method converges for $\hat{\tau} = 410.427 \times 10^{-15}$ s, $\hat{a} = 0.982$, and Newton's method for $\hat{\tau} = 409.684 \times 10^{-15}$ s, $\hat{a} = 0.984$, i.e. the value obtained by the gradient method is 0.18% greater than that obtained through Newton's method. The fact that these two completely different methods yield results close to one another is a good validation for both algorithms. The minimum calculated using Newton's method will be used for the rest of this analysis.



(a) Minimisation of the 2D NLL.

(b) Convergence of minimisation algorithms.

Figure 5: Minimisation of the 2D NLL function. Figure (a) shows a contour plot of the 2d NLL function around the minimum, overlaid with the path traced by the gradient method (green) and Newton's method (red). Figure (b) shows the convergence of the two algorithms.

The error on the 2 dimensional NLL estimate is calculated using the method described in section 2.1, using the Newton-Raphson method repeatedly on slices of the NLL function, for a range of a values near the minimum. The algorithm struggles to converge near $\sigma_a^{\pm}$, as this corresponds to a repeated root with a derivative close to zero. However, as the main aim is to calculate the $\sigma_\tau^{\pm}$ this loss in accuracy is acceptable, and a reasonably accurate value is found by increasing the density of slices. The resulting ellipse is shown in figure 6 for 1000 slices across a, using the estimate derived using Newton's method for $\hat{\tau}$. The error can be shown to be the outermost points of the ellipse

(Barlow 1989). This yields a final 2 dimensional NLL estimate of $\hat{\tau} = (409.684^{+5.555}_{-5.436}) \times 10^{-15}$ s and $\hat{a} = (0.984^{+0.008}_{-0.009}) \times 10^{-15}$ s. This is a percentage error of 0.101% from the value quoted by Olive et. al. (2014), and corresponds to a percentage difference of 1.25% from the 1-dimensional NLL estimate for $\hat{\tau}$. It should be noted that the error on the lifetime increases in the 2 dimensional case compared to the 1 dimensional case, which makes sense as an extra parameter has been introduced with the same amount of data available to constrain both of them.

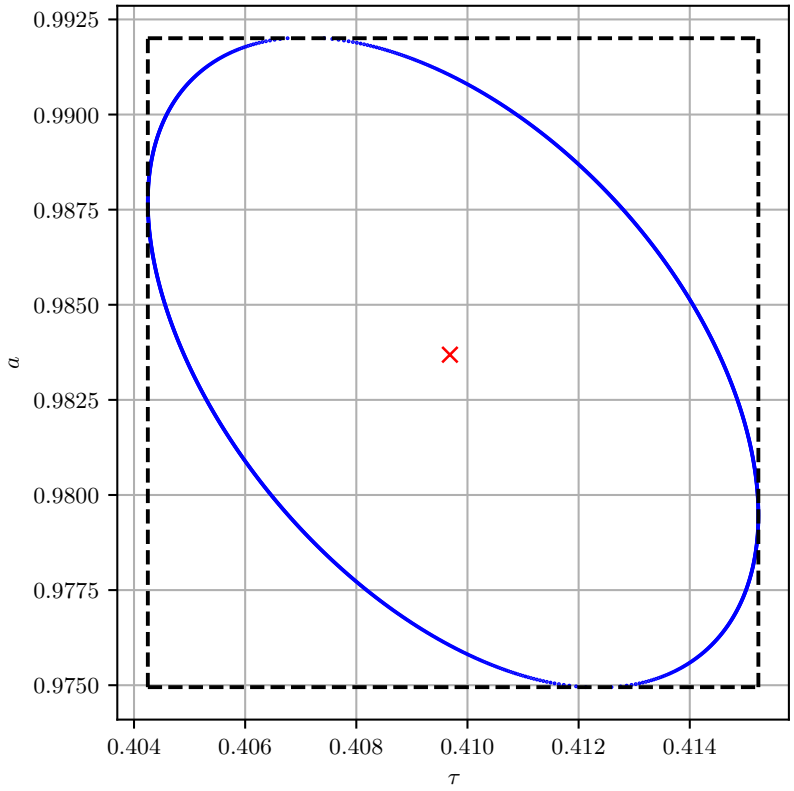


Figure 6: Contour line satisfying $\mathrm{NLL}(\tau, a) = \mathrm{NLL}(\hat{\tau}, \hat{a}) + 1/2$ (blue line), using $\hat{\tau}$ from Newton's method (red cross), with lines corresponding to $\sigma^{\pm}_{\tau}$ and $\sigma^{\pm}_{a}$ (black dotted line).

4 Conclusion

In this report, the lifetime τ of a D^0 meson has been estimated from a data set of measured lifetimes and uncertainties, using NLL estimation. The data was initially fitted to a 1 dimensional distribution of measured lifetimes, represented by an exponential distribution convoluted with a Gaussian. Minimising the resulting NLL function using parabolic minimisation yielded an estimate $\hat{\tau} = (404.546^{+4.738}_{-4.675}) \times 10^{-15}$ s, which is a percentage error of 1.35% from the value quoted by the Particle Data Group (Olive et. al. 2014). The error on this estimate was calculated using two methods, and the variation of the error with data set size was investigated through a linear regression, obtaining the standard result that the variance of the mean scales as $\sigma_{\overline{x}} \propto 1/\sqrt{N}$. The distribution was then made more realistic by introducing a parameter a to represent the proportion of true signal to background signal in the data set. The 2 dimensional NLL function obtained from this distribution was minimised for $(\hat{\tau}, \hat{a})$ using the gradient method and Newton's method,

obtaining similar values in both cases (a percentage difference of 0.18%). The values obtained through Newton's method are: $\hat{\tau} = (409.684^{+5.555}_{-5.436}) \times 10^{-15}$ s and $\hat{a} = (0.984^{+0.008}_{-0.009}) \times 10^{-15}$ s, representing a percentage error of 0.101% from Olive et. al. (2014) and a percentage difference of 1.25% from the 1 dimensional NLL estimate. In both the 1 and 2 dimensional cases, the values of $\hat{\tau}$ fall within the range allowed by the PDG value, taking into account uncertainties on both values.

Possible improvements to the code include implementing adaptive step sizes in order to reduce the number of steps required in the 2 dimensional case, starting with a greater step size far from the minimum and reducing the size as the minimum is approached. Although execution time was not an issue in this specific case, it was still necessary to optimise the step size for the gradient method manually which could be more time-consuming for more complicated functions, in particular with multiple minima.

References

Barlow, R. (1989), *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Wiley.

Griffiths, D. J. (2008), *Introduction to elementary particles; 2nd rev. version*, Physics textbook, Wiley, New York, NY.
URL: *https://cds.cern.ch/record/111880*

Olive et. al., K. (2014), 'Review of particle physics', *Chin. Phys. C* **38**(9), 978.
URL: *https://doi.org/10.1088/1674-1137/38/9/090001*

Press, W. (2002), *Numerical Recipes in C++: The Art of Scientific Computing*, Cambridge University Press.

Weisstein, E. W. (2017), 'Newton's method'. [Online; accessed December 10, 2017].
URL: *https://mathworld.wolfram.com/NewtonsMethod.html*