Department of Physics & Astronomy
Ithaca College

pydls

A Python Suite for the Bayesian Inferrence of Particle Size Distributions in Dynamic Light Scattering Experiments

Thy Doan Mai Le
Advisor: Jerome Fung

Date: September 30, 2019

Contents

1 Dynamic Light Scattering

1.1 What is Dynamic Light Scattering?

Dynamic light scattering is a useful technique for understanding and studying the characteristics and properties of soft matter systems and colloidal samples, solid particles suspended in some substrate. In order to study these particles, light from the visible spectrum is directed at the colloidal sample and upon encountering the particles within the sample, the incident light scatters. A photomultiplier located at the other end of the setup will detect any of the scattered photons that leave the colloidal sample. Each time a photon is detected by the photomultiplier, a voltage spike with amplitude proportional to the detected photon's intensity is generated. Because the particles within the colloidal sample are also subject to random Brownian motion over time, the intensity of scattered light is expected to fluctuate over time, resulting in a time-dependent fluctuating signal. To extract useful information from a fluctuating signal, the detector's computer system will use the raw signal to generate an autocorrelation function that encapsulates information regarding the kinetics, the shapes and sizes of the particles that scattered the incident light. A general setup of a dynamic light scattering experiment can be found in Fig. 1.

1.2 Autocorrelation Function

A correlation function describes how similar two properties are. An autocorrelation function, then, shows how similar a property is to itself at different points in time. The autocorrelation function of a function $f(t)$ is then:

$$\langle f(t)f(\tau)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T f(t)f(t+\tau)dt \tag{1}$$

where τ is some interval of time between two points of comparison. At $t = 0$, the function is exactly itself and therefore the autocorrelation has a maximum at $t = 0$. As some time passes, if the function begins to have random fluctuations, the autocorrelation value will begin to drop. If the fluctuations become very invariant and drastic, then the autocorrelation value will drop to zero because at this point in time, the function does not resemble its original shape anymore. An example of this drop in autocorrelation value is demonstrated in Fig. 2

1.3 Project Goals

A recent study published in 2014 by Boualem, Jabloun, Ravier, Naiim and Jalocha[**?**] proposed the use of an MCMC (Monte Carlo Markov chain) algorithm for Bayesian inferrence of particle size distribution from dynamic light scattering experiments. The publication noted successfully tested theoretical models along with simulation test results, but left out any references to the actual program that was used to generate said results. An algorithm to determine particle size distribution using Bayesian inference would be a powerful tool for any researcher studying soft matter systems. Hence, this project attempts to replicate the algorithm that was proposed in the aforementioned study. If successfully tested, the ultimate goal of this project would be to create a Python package for Bayesian inference of the particle size distribution from any raw DLS data of a single angle, and publish the package in an open-source platform for all to utilize.

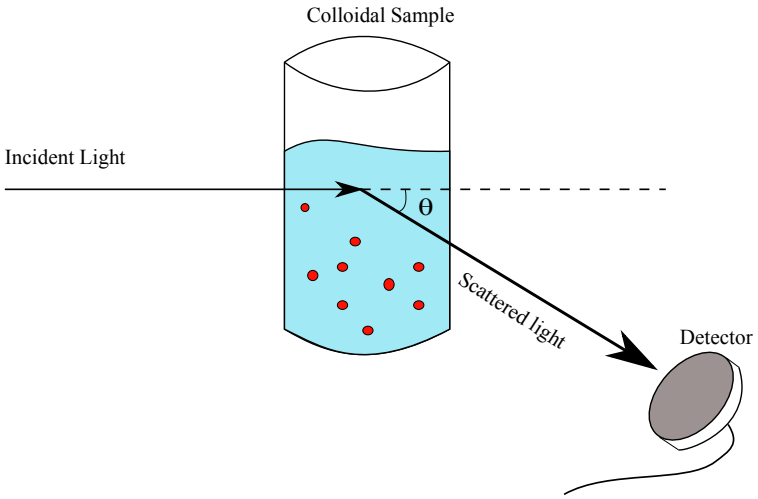

Figure 1: When incident light encounters the colloidal sample, it is scattered by each individual particulate molecule within the sample and upon exiting the sample, all individually scattered light rays come together to give an interference pattern that is then received by the detector. Because the individual particles within the sample all experience random Brownian motion, the interference pattern is predicted to have a fluctuating intensity as the molecules move into and out of the incident light beam.

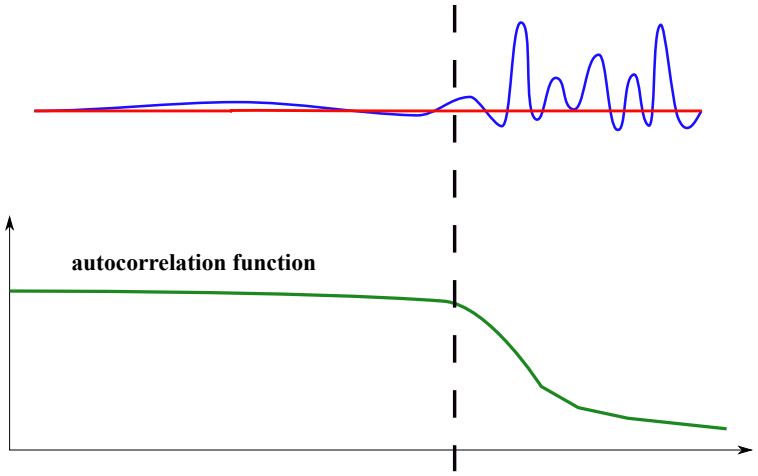

Figure 2: The red line describe the function's initial state, projected onto the time axis for comparison with the blue line, the function's behavior at later times. The autocorrelation function in green begins at 1 but slowly drops to almost 0 as the function's behavior becomes more erratic.

2 Theoretical Background

2.1 Prior Distribution

According to Buoalem, Jabloun, Ravier, Naiim and Jalocha, the prior distribution of this non-parametrized method is as follows:

$$p(\mathbf{f}) \propto exp\left(\mathbf{f}^T {\mathbf{L_2}}^T \mathbf{L_2 f}\right), \text{if } \mathbf{g} \geq 0 \tag{2}$$

where $\mathbf{f}^T$ stands for the transpose of the input function, ${\mathbf{L_2}}^T$ stands for the transpose of the second derivative operator and $\mathbf{L_2}$ stands for the second derivative operator. This prior distribution aims to penalize any curvature in the posterior distribution so as to avoid inaccurate maxima.

2.2 Likelihood Distribution

The likelihood distribution is modeled according to the $g^{(2)}$ function as

$$p(\tilde{\mathbf{g}}^{(\mathbf{2})}|\mathbf{f}, \sigma_r^2) = \frac{1}{(2\pi)^{\frac{M_r}{2}} \sigma^{M_r}} exp\left(-\frac{\chi_r(\mathbf{f})}{2\sigma_r^2}\right) \tag{3}$$

if $\mathbf{f} \geq 0$.

The χ-squared $\chi_r(\mathbf{f})$ is defined as

$$\chi_r(\mathbf{f}) = \sum_{j=1}^{M_r} \left(\tilde{g}_{\theta_r}^{(2)}(\tau_j) - g_{\theta_r}^{(2)}(\tau_j)\right)^2 \tag{4}$$

2.3 Posterior Distribution

With the prior and likelihood distributions described in Eq. 2 and Eq. 3, the posterior distribution for the particle size distribution with the noise variance given the measured autocorrelation data ($g^{(2)}$) is:

$$p(\mathbf{f}, \sigma_1^2, \ldots, \sigma_r^2|\tilde{\mathbf{g}}^{(\mathbf{2})}) \propto exp\left(-\mathbf{f}^T {\mathbf{L_2}}^T \mathbf{L_2}\right) \prod_{r=1}^{R} \left[\frac{1}{\sigma_r^{M_r+2}} exp\left(-\frac{\chi_r(\mathbf{f})}{2\sigma_r^2}\right)\right] \tag{5}$$

Using the classical Jeffrey prior density for the noise variance in $\tilde{\mathbf{g}}^{(\mathbf{2})}$,

$$p(\sigma_r^2) \propto \frac{1}{\sigma_r^2} \tag{6}$$

the posterior distribution can be integrated over the noise variance with the following integral:

$$\int_0^\infty \frac{1}{\sigma_r^{M_r+2}} exp\left(-\frac{\chi_r(\mathbf{f})}{2\sigma_r^2}\right) d\sigma_r \tag{7}$$

In order to integrate this integral, let $\sigma_r = \frac{1}{t}$ and thus $d\sigma_r = -\frac{1}{t^2}dt$. From this, we can rewrite expressions from the integrand in Eq. 7 as follows.

$$\begin{aligned} \frac{1}{\sigma_R^{M_r+2}} &= t^{M_r+2} \\ \frac{\chi_r(\mathbf{f})}{2\sigma_r^2} &= \frac{\chi_r(\mathbf{f})}{2}t^2 \end{aligned} \tag{8}$$

Thus the integral from Eq. 7 becomes:

$$\begin{aligned} &\int_{\infty}^{0} t^{M_r+2}\left(-\frac{1}{t^2}\right) exp\left(-\frac{\chi_r(\mathbf{f})}{2}t^2\right) dt \\ &= \int_{0}^{\infty} t^{M_r} exp\left(-\frac{\chi_r(\mathbf{f})}{2}t^2\right) dt \end{aligned} \tag{9}$$

The limits switched from $(0,\ \infty)$ in Eq. 7 to $(\infty,\ 0)$ in Eq. 9 because of the substitution $\sigma_r = \frac{1}{t}$. The reciprocal relationship of the substitution switched the limits of the integration. In order to make the integral easier to integrate, the minus sign on the second line of Eq. 9 is exchanged for the re-switching of the integration limits from $(\infty,\ 0)$ back to $(0,\ \infty)$.

In order to actually integrate Eq. 9, a second substitution has to be made. Let $\tau = t\sqrt{\chi_r}$ and $\mathrm{d}\tau = \sqrt{\tau}\mathrm{dt}$. From this, we can rewrite the expressions in the integrand of Eq. 9 as follows.

$$\begin{aligned} \chi_r(\mathbf{f})t^2 &= \tau^2 \\ t^{M_r} &= \left(\frac{\tau}{\sqrt{\chi_r}}\right)_r^M \\ &= \tau^{M_r}\chi_r^{-\frac{M_r}{2}} \end{aligned} \tag{10}$$

From the above substitutions, the integral becomes:

$$\begin{aligned} &\int_{0}^{\infty} \chi_r^{-\frac{1}{2}}\tau^{M_r}\chi_r^{-\frac{M_r}{2}} exp\left(-\frac{\tau^2}{2}\right) d\tau \\ &= \int_{0}^{\infty} \chi_r^{-\frac{1}{2}(M_r+1)}\tau^{M_r} exp\left(-\frac{\tau^2}{2}\right) d\tau \end{aligned} \tag{11}$$

Since the χ_r term does not have τ dependence, we can pull that out of the integrand, giving us the following integral at the end:

$$\chi_r^{-\frac{1}{2}(M_r+1)} \int_{0}^{\infty} \tau^{M_r} exp\left(-\frac{\tau^2}{2}\right) d\tau \tag{12}$$

Recall that the χ_r expression was previously

$$\chi_r(\mathbf{f}) = \sum_{j=1}^{M_r} \left(\tilde{g}_{\theta_r}^{(2)}(\tau_j) - g_{\theta_r}^{(2)}(\tau_j)\right)^2 \tag{13}$$

where $\tilde{g}^{(2)}$ is modeled to be:

$$\begin{aligned} \tilde{g}^{(2)}_{\theta_r}(\tau_j) &= 1 + \beta_r \left(k_{\theta_r} \sum_{i=1}^{N} f(D_i) C_{I,\theta_r}(D_i) exp\left(-\frac{\Gamma_{0,\theta_r}}{D} \tau_j \right) \Delta D_i \right)^2 \\ \text{where } \Gamma_{0,\theta_r} &= \frac{16\pi n^2 \sin^2(\theta/2) k_B T}{2\lambda_0^2 \eta} \end{aligned} \tag{14}$$

with

$$\begin{aligned} &\text{n} \ : \text{refractive index} \\ &\theta \ : \text{angle at which scattered light was measured} \\ &k_B \ : \text{Boltzmann constant} \\ &T \ : \text{temperature in Kelvin} \\ &\lambda_0 \ : \text{incident wavelength value in vacuum} \\ &\eta \ : \text{viscosity} \\ &D_i \ : \text{diameter of particle size} \\ &\beta_r \ : \text{instrumental factor} \\ &k_{\theta_r} \ : \text{normalization constant that can be found via k} = \frac{1}{\int_0^\infty f(D) C_{I,\theta}(D) dD} \\ &\tau_j \ : \text{delay times} \\ &f(D) \ : \text{the particle size distribution function} \end{aligned} \tag{15}$$

The posterior distribution is now defined to be:

$$p(\mathbf{f}|\tilde{\mathbf{g}}^{(2)}) \propto exp\left(-\mathbf{f}^T \mathbf{L_2}^T \mathbf{L_2} \mathbf{f}\right) \prod_{r=1}^{R} \left[\chi_r(\mathbf{f})\right]^{-\frac{M_r}{2}} \tag{16}$$

Since $\chi_r(\mathbf{f}) = \sum \left(\tilde{g}^{(2)}_{\theta_r}(\tau_j) - g^{(2)}_{\theta_r}(\tau_j) \right)^2$, the posterior distribution is essentially:

$$p(\mathbf{f}|\tilde{\mathbf{g}}^{(2)}) \propto exp\left(-\mathbf{f}^T \mathbf{L_2}^T \mathbf{L_2} \mathbf{f}\right) \prod_{r=1}^{R} \left(\sum \left(\tilde{g}^{(2)}_{\theta_r}(\tau_j) - g^{(2)}_{\theta_r}(\tau_j) \right)^2 \right)^{-\frac{M_r}{2}} \tag{17}$$

When the log of the posterior is taken, the above equation becomes:

$$\ln(p) = -\mathbf{f}^T \mathbf{L_2}^T \mathbf{L_2} \mathbf{f} + \left(-\frac{M_r}{2} \right) \ln \left(\sum \left(\tilde{g}^{(2)}_{\theta_r}(\tau_j) - g^{(2)}_{\theta_r}(\tau_j) \right)^2 \right) \tag{18}$$

2.4 Development

Testing and development records can be found at `https://github.com/thydmle/pydls`

3 Conclusion

Comprehensive testing showed that pydls is successfully able to handle DLS autocorrelation functions from single angle data with very high accuracy. Further development should target the expansion of this package into a code base that is able to handle multi-angle DLS data with accurate (instead of an approximation) Mie scattering coefficient calculations.