Notes on Fitting and Analysis Frameworks for QENS Spectra of (Soft) Colloid Suspensions

Christian Beck[1,2,*], *Kevin* Pounot[1,2,**], *Ilaria* Mosca[1,2], *Niina* H Jalarvo[3,4], *Felix* Roosen-Runge[5], *Frank* Schreiber[1], and *Tilo* Seydel[2,***]

[1]Institut für Angewandte Physik, Universität Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany.
[2]Institut Max von Laue - Paul Langevin (ILL), CS 20156, F-38042 Grenoble Cedex 9, France.
[3]Jülich Centre for Neutron Science (JCNS), Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany
[4]Chemical and Engineering Materials Division, Neutron Sciences Directorate, and JCNS Outstation at the Spallation Neutron Source (SNS), Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831, USA
[5]Department of Biomedical Sciences and Biofilms-Research Center for Biointerfaces (BRCB), Malmö University, 20506 Malmö, Sweden.

Abstract. With continuously improving signal-to-noise ratios, a statistically sound analysis of quasi-elastic neutron scattering (QENS) spectra requires to fit increasingly complex models which poses several challenges. Simultaneous fits of the spectra for all recorded values of the momentum transfer become a standard approach. Spectrometers at spallation sources can have a complicated non-Gaussian resolution function which has to be described most accurately. At the same time, to speed up the fitting, an analytical convolution with this resolution function is of interest. Here, we discuss basic concepts to efficient approaches for fits of QENS spectra based on standard MATLAB and Python fit algorithms. We illustrate the fits with example data from IN16B, BASIS, and BATS.

1 Introduction

The interpretation of QENS data is generally performed in terms of maximum likelihood methods [1], commonly known as “data fitting”. By this curve fitting, the most probable from an infinite set of hypotheses expressed by a continuous variable, namely the fit parameter, is determined by finding the optimum of a suitable likelihood function [1–3]. For the generally assumed situation of Poisson-distributed data points, the maximum likelihood method is mathematically equivalent to the least-squares method [1], and thus, QENS data fitting is generally based on least-squares minimization. The data fitting procedure also involves the determination of the confidence bounds on the fit parameters as well as of the goodness of fit.

Along with the higher brightness and better signal-to-noise ratios of recent neutron spectrometers [4], the topic of associated data analysis software frameworks is of continuous interest [5–7]. Better scattering signals allow for more complex models to be tested. The switching between numerous models to be applied on a set of data has to be quick and easy. Here, script-based approaches seem to be unavoidable, and concepts for easily modifiable scripts gain interest. Models of QENS data routinely contain several Lorentzians that are coupled in intensity and momentum transfer [8]. An example consists in the model of coupled diffusive states that for instance describes the internal diffusion of proteins, accounting for superimposed protein backbone and side chain fluctuations [9].

The model is convoluted with the resolution function of the instrument - obtained on dedicated data - to describe a signal that can be fitted to the experimental data. In cases where the model consists only of Lorentzians, Gaussians, and elastic contributions or a sum thereof, an analytical convolution can be readily performed with the benefit of low computational cost if the spectrometer resolution function is described by a sum of an arbitrarily large number of Gaussian (and Lorentzian) functions. In this case, the scattering function simply consists of a sum of Voigt (and Lorentzian) functions. The Fourier transform of the Mittag-Leffler function [10, 11], namely the generalized Lorentzian function, constitutes another important function for fitting of QENS spectra, which contains the stretched exponential. This function cannot yet be readily included in a generalized Voigt function for an analytical convolution, although initial approaches to generalize the Voigt function seem to exist in the mathematical literature [12], and Padé-approximations can be applied near the elastic line [13]. Fourier transform methods can be used alternatively, but are subject to the well-known general limitations and maybe more difficult to apply for non-symmetric resolution functions [14, 15].

The present article addresses script concepts to test different models for the scattering function efficiently. It

[*] e-mail: christian.beck@uni-tuebingen.de
[**] current address: European Synchrotron Radiation Facility - 71, avenue des Martyrs, CS 40220, 38043 Grenoble Cedex 9, France.
[***] e-mail: seydel@ill.eu

does not, however, address Bayesian inference, for instance regarding the number of Lorentzians present in a data set. This topic has been dealt with previously by other work [16, 17]. Moreover, it is noted that several standard procedures for fitting are available inside standard software such as Mantid [18].

Subsequent to a brief outline of script concepts in MATLAB and Python, this article illustrates the analytical convolution with the spectrometer resolution function employing IN16B, BASIS, and BATS example data on protein solutions. The aim of this work is to show the general applicability of the analysis framework to different instruments but not the comparison of instruments. Therefore, different spectra of similar, but not identical samples are analyzed.

2 Analysis

The analysis of the collected QENS spectra often implies a separation of different contributions to the scattering signal from superimposed motions as well as from immobile fractions. The contribution of the sample holder can be removed based on a separate measurement. For the subtraction of its signal from other spectra, the angular dependence of absorption can be accounted for using methods such as the Paalman-Pings corrections [19] . Additionally, the contribution of some sample components can be reduced by (partial) deuteration (i.e., using D_2O as solvent). However, it should be mentioned that the deuteration can significantly influence the system investigated [20]. Similarly to the sample holder, the solvent can also be measured individually. However, depending on the solute, the solvent composition might change due to effects such as dissolving ions or sugars and H-D exchange. In addition, the solvent contribution has to be rescaled based on the excluded volume by the solute [21]. The reduced solvent contribution can either be subtracted from the measured spectra or can be included in the model as fixed contribution. For the later option, a corresponding model for the solvent has to be applied, which might also depend on the energy and momentum transfer range probed [21].

The signal from dissolved soft ("floppy") colloids such as proteins consists of several different components. While a translational and rotational center-of-mass diffusion [21] can be observed due to the nature of the dissolved objects, also internal fluctuations, such as protein backbone and side chain fluctuations, and mesoscopic domain motions such as lobe motions in an antibody protein might be observed [8, 22, 23]. In order to separate these different contributions, an approach with relatively few assumptions based on fitting each spectrum for each momentum transfer q individually, may not be sufficient. A simultaneous fit approach, taking into account both energy and momentum transfer, is then necessary. A detailed description of the procedure, along with an interpretation of the models, can be found in the corresponding literature [8].

Given the recent technical improvements of the instruments allowing to access high energy transfers with an improved energy resolution, it becomes more important to have the possibility to fit more complex models, since the energy or time scales investigated are increased and diffusive contributions, which were masked previously either by the broad resolution function or by the small energy range investigated, become visible. Such a global fit offers on the one hand the possibility to extract more reliable parameters; on the other hand, it requires a higher level of modelling. Since several parameters, such as the parametrization of the resolution or the solvent contribution depend on q, a global fit needs to have q-dependent fixed parameters, as well as q-independent and q-dependent fit parameters. In the event that not only one single spectrum, but several spectra are analyzed simultaneously, e.g., to extract temperature independent entropy and enthalpy values instead of free energies [24], the complexity of the model increases even further, since the background signal dominated by the solvent is also temperature dependent.

The complexity of the system can also be increased by crowding through the addition of macromolecules in the aqueous suspension. Not only the solvent, but also parts of the crowders can be deuterated to mimic more complex environments, such as living cells, and to obtain a scattering signal which is dominated by the contribution of the tracer particle of interest [25–27].

Similarly to approaches already employed in small angle scattering data analysis [28, 29], the separation can be performed based on the knowledge of the sample composition. Since the signal is typically dominated by incoherent scattering, the q-dependence mainly influences the different dynamic features. While this approach includes the separation of the (apparent) global diffusion from internal diffusive processes [8], the same procedure can be applied to separate the scattering signal of the crowders from the one of the protein of interest [23], the contributions of monomers and crystals [30], the contributions of two different proteins [31], or to distinguish the scattering signal of proteins from the one coming from the detergent used for purification [32].

3 "Wrapping" in Python for flexible models

To keep the readability of the analysis code as good as possible as well as generally applicable, it is important to be able to easily change the model for the resolution function to apply the code to different spectrometers. This flexibility implies an automated convolution of the model function for the samples with the resolution function. Additionally, a change between different models can easily be performed in this way. Without going into any further detail of coding, such a flexible fit function call can for instance be written as follows using *python* and *scipy.curve_fit*:

```
popt, pcov = curve_fit( lambda x, *p: \
wrapper_function( x, q, n, r, len( f0 ), \
p, **keyword_parameters ), \
x, y, p0=f0, bounds=(l,u), sigma=dy )
```

Therein, the concept of the *lambda* function allows to point to external functions, and the *wrapper_function* contains the fit function that can be modified by a choice from a library of functions or a building-block syntax for functions. This choice from a library or selection of components is passed via suitable keyword arguments ***keyword_parameters*. These keyword arguments can take the form of an arbitrarily long list contained for instance in a *python* dictionary (*dict*). Besides the energy axis x, the q values and the used indices n on these q values can be passed as vectors. Moreover, the matrix r containing the parameters defining the analytical resolution function can be provided. $f0$ denotes the initial guess, i.e., the start values of the fit, and l and u the lower and upper bounds, respectively, and dy the standard errors on the spectral signal y. Importantly, the vector of fit parameters p can have a variable length, which is identical to the length of the initial guess *len(f0)*. The fit returns the optimum parameters *popt* and the covariance matrix *pcov*. The *wrapper_function* with the interface

```
def wrapper_function( x, q, n, r, N, \
*args, **kwargs ):
    ...
```

extracts the fit parameters into a vector *fp* usable subsequently via

```
a  = list( args[0][:N] )
fp = np.asarray( a )
```

Similar or technically simpler coding solutions can be achieved in other software such as MATLAB, as briefly outlined in the following section.

4 Fits in MATLAB

In MATLAB, simultaneous fits for all q at once to the data y at the points in energy x can be performed using the *lsqnonlin* command from the optimization toolbox:

```
lsqnonlin(model,f0,l,u,options,x,y);
```

Therein, as in the *python* version, $f0$ is the initial guess, and l and u are the lower and upper bounds, respectively. Importantly, x and y can be matrices. *model* is a function handle pointing to the function *EvaluatedModel* which returns a function of the model evaluated for the parameters fp. This "function of the model" has to take care of the weighting by the errors dy on the data, since *lsqnonlin* does not provide an option for explicit error weighting:

```
model=@(fp, xdata, ydata)EvaluatedModel(...
dy, q, fp, xdata, ydata, N, fixedp);
```

This syntax establishes that *EvaluatedModel* can take more parameters – namely all parameters in the second pair of round brackets – than the function handle *model* is allowed to accept – namely only the parameters in the first pair of round brackets – and therefore allows to hand over the number of q-independent fit parameters N as well as the fixed parameters and resolution parameters *fixedp*, and the q values, to calculate the scattering function depending on q and $\hbar\omega$. Hence, *model* in this MATLAB implementation takes the equivalent role of the *wrapper_function* outlined in the preceding section describing the *python* implementation, and the function handle is defined by the "@" symbol in MATLAB as opposed to the *lambda* function in *python*. The fit algorithm performs $\min_{fp} \|f(fp)\|_2^2$ with the fit parameters *fp* and, thus, represents a least-squares minimization.

Table 1: Specifications of the spectra shown in Fig. 1.

Subfigure	a	b	c
Instrument	IN16b	BASIS	BATS
number of q values/groups	18	18	19
number of $\hbar\omega$ values	1024	500	4126
mean spacing q [Å^{-1}]	0.096	0.010	0.094
mean spacing $\hbar\omega$ [μeV]	0.058	0.4	0.1
sample can outer ⌀[mm]	23	23	14

5 Analytical convolution

In Figure 1, we show the resolution functions for the different NBS Spectrometers IN16b (ILL) in its standard energy resolution high flux mode with Si(111) analyzers, of BASIS (SNS), and of IN16b in its BATS option (using the resolution setting lres4). All the resolution functions $\mathscr{R}$ were approximated with a sum of several Gaussian functions $\mathscr{G}(\sigma, x)$:

$$\mathscr{R} = \sum_n a_n \mathscr{G}\left(\sigma_n, \omega - \omega_n\right) \tag{1}$$

using $n = 2$ for the different options of IN16b and $n = 5$ for the BASIS resolution and a flat background with a_n, σ_n and ω_n being q-dependent scaling parameters, variances and center positions of the Gaussian function n, respectively. These parameters might depend on the binning applied to the q and $\hbar\omega$ space.

Any model of a scattering function containing only Lorentzians, including an elastic contribution, Gaussians, and Voigt functions, can be easily represented by an analytical convolution with these resolution functions.

Conveniently, the Voigt function $\mathscr{V}$, as the convolution of Gaussian with Lorentzian function, is available in *python* via the Faddeeva function *scipy.special.wofz*[33], $f(z) = \exp(-z**2)*\mathrm{erfc}(-i*z)$:

```
import numpy as np
from scipy.special import wofz

def V( x, sigma, gamma ):
return np.real( wofz( (x + 1j*gamma) \
      / sigma / np.sqrt(2.) ) )      \
      / sigma / np.sqrt( 2.*np.pi )
```

The voigt function $\mathscr{V}$ also exists as *scipy.special.voigt_profile* since *scipy* 1.9.1.

In MATLAB, the Voigt function can be implemented as C++ code – for instance via the Faddeeva package available from *mathworks.com* – and linked as compiled *mex* file.

In Figure 2, the spectra of different aqueous (D_2O) solutions containing the dissolved protein *bovine serum*

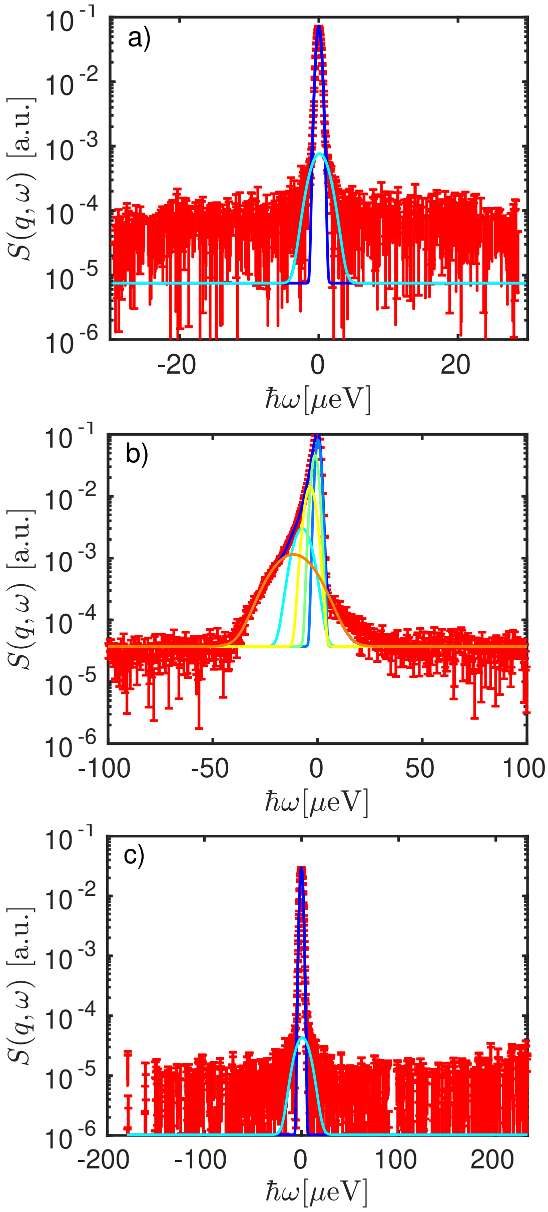

Figure 1: Resolution functions of different instruments (red symbols) and corresponding fits with the different Gaussian functions (colored lines). Different vanadium samples were measured with different statistics as well as different geometries, and different q and $\hbar\omega$ binnings are applied. A direct comparison between the instruments is therefore not possible. In contrast, the present and subsequent figure illustrate the generality of the analysis approach. For the different instruments, a different number n of Gaussian functions was used in addition to a flat background. a) IN16b: $n = 2$; $q = 1.03$ Å^{-1} b) BASIS: $n = 5$; $q = 1.05$ Å^{-1} c) BATS: $n = 2$; $q = 1.03$ Å^{-1}. Details of the binning applied are given in Table 1.

albumin (BSA) at concentrations between $c_p = 100 - 200$ mg/ml are shown which are measured at the temperatures T=280 K and T=295 K. A direct comparison of the spectra is not possible due to the temperature dependence [22, 34] and the volume fraction scaling [34]. A model using two Lorentzian functions $\mathscr{L}$ to describe the apparent global center-of-mass diffusion coefficient and to summarize the internal diffusive processes of the proteins, which is established for different proteins in the small energy transfer range [8, 23, 25, 30, 35], and one Lorentzian function for the solvent contribution is convoluted with the resolution function to obtain the final fit model

$$\begin{aligned} S(q,\omega) &= \mathscr{R} \otimes \Big(A_0 \cdot \mathscr{L}_\gamma(\omega) + \\ &\quad (1-A_0)\cdot \mathscr{L}_{\gamma+\Gamma}(\omega) + \beta_{D_2O} \cdot \mathscr{L}_{\gamma_{D_2O}}(\omega)\Big) \qquad (2) \\ &= A_0 \sum_n a_n \mathscr{V}_{\sigma_n}(\gamma, \omega-\omega_n) + \\ &\quad (1-A_0)\sum_n a_n \mathscr{V}_{\sigma_n}(\gamma+\Gamma, \omega-\omega_n) + \\ &\quad \beta_{D_2O}\sum_n a_n \mathscr{V}_{\sigma_n}(\gamma_{D_2O}, \omega-\omega_n) \qquad (3) \end{aligned}$$

with a_n, σ_n, ω_n being the q dependent parameters fixed by the resolution function from Equation 1, $\gamma_{\mathrm{D_2O}}, \beta_{\mathrm{D_2O}}$ the parameters fixed based on the measurements of the solvent. For the apparent global diffusion, the individual contributions due to the different Gaussian functions used for the resolution function are displayed, while the internal diffusive processes and the solvent contribution are shown as a sum of the corresponding Voigt functions. By investigating the different contributions, it becomes visible that especially in the case of non-symmetrical resolution functions, it is important to consider the additional contributions. It should be mentioned that the internal diffusive processes are better accounted for by two coupled Lorentzian functions describing the motion of the protein backbone and side chains individually [22, 24]. For reasons of comparability, the same model is applied for the different energy ranges. The spectra shown illustrate the applicability of the approach for spectra from spectrometers located at spallation sources such as BASIS and also for spectrometers at continuous sources (IN16b, BATS) which are often characterized by more symmetrical resolution functions.

6 Global optimization

The complexity of the models used to fit the experimental data implies that the fit function is not necessarily strictly convex, but multiple local minima can be found. Hence, the least-square algorithm is sensitive to the initial values for the parameters given by the user. To circumvent this problem, common sense and experience can help discarding solutions for which the fitted parameters are aberrant. However, multiple solutions can make sense in regard to the system studied and more robust methods are needed to find the different minima and identify the global minimum. To this end, global optimization methods can be used, such as the genetic algorithm [36] or basin-hopping [37]. The basin-hopping algorithm consists in running

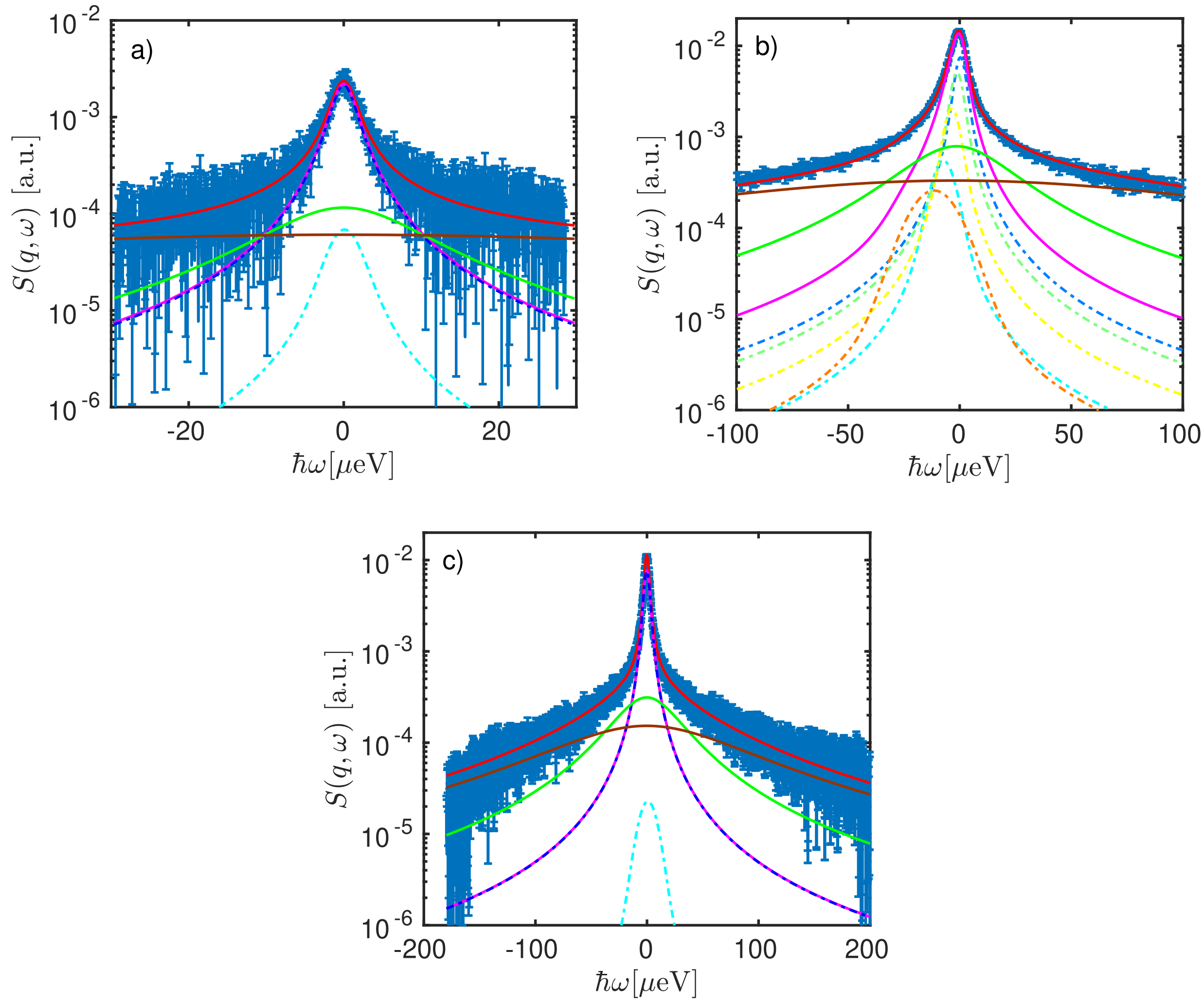

Figure 2: Spectra from aqueous (D_2O) BSA protein solutions recorded on different spectrometers (symbols) and associated fits (solid red lines superimposed on the symbols). The individual Voigt components associated with the convoluted apparent center-of-mass diffusion (magenta line) corresponding to each Gaussian in the resolution model are displayed as dash-dotted lines, with the colors corresponding to the ones used in Figure 1. In addition, the rescaled water contribution and the internal diffusion contribution are depicted as solid brown and green lines, respectively, without displaying their individual Voigt components. The protein concentrations in the samples as well as the sample temperatures differ for the spectra displayed: a) IN16b: c_p = 100 mg/ml; T = 280 K; q = 1.0 Å^{-1}; b) BASIS: c_p = 200 mg/ml; T = 295 K; q = 1.05 Å^{-1}; c) BATS: c_p = 150 mg/ml; T = 280 K; q = 1.05 Å^{-1}. Details of the different binning applied are given in Table 1.

several iterations of least-square minimization where the initial parameters are randomly updated. The new solution is typically accepted or discarded using the so-called Metropolis criterion from the Monte-Carlo method [38]. Numerous global optimization algorithms are available in *Python* via the *scipy.optimize* module. It is noteworthy that the signature and return types of these functions differ from the *curve_fit* function described above. For instance, the *basin-hopping* function call will read:

```
from scipy.optimize import basinhopping

result = basinhopping(
    fit_function,
    x0,
    minimizer_kwargs={
        args: additional_arguments
    }
)

optimal_params = result.x
```

where *x0* is a list of initial parameters for the first iteration and the *fit_function* has the signature (using Python type hints):

```
def fit_function(
        p: list[float],
        *additional_arguments) -> float:
    # compute the cost function
    # (can be chi-square function or other)
    cost = ...
    return cost
```

where *additional_arguments* can contain p, x, q and r as described above.

7 Future developments

In case the relevant diffusive processes of the sample investigated cannot be covered by one single instrument, different instruments or instrumental setups can be used to access different time and length scales. As a first approach, the corresponding spectra can be analyzed individually and the corresponding results can be further analyzed with models covering the different time scales [32]. In case it is desired to perform fits combining different spectrometers, it is necessary to employ frameworks which not only take into account the specific quantities for different instruments, but which also use a weighting between the different instruments based on data quality, instrument flux and amount of collected data points. A correct treatment of the error propagation is even more crucial in this case.

8 Conclusions

We have presented conceptual notes on MATLAB and Python implementations allowing for efficient scripting and fast fitting of neutron backscattering data, due to an analytical convolution. We have illustrated this fit procedure on example protein solution spectra. Spectrometer-design specific resolution functions can be routinely handled even in the case of strongly non-Gaussian resolution functions. Due to the computing time-saving analytical convolution, rather complex fits, including simultaneous fits of the spectra at all momentum transfers at once, can be carried out rapidly.

9 Code and Data Availability

Matlab analysis code performing multidimensional fits for individual and several spectra are available from the corresponding author on reasonable request. python code is available from *https://github.com/seydel/QENS_utilities* as well as from *https://github.com/kpounot/nPDyn*. The experimental data collected at IN16b are curated by the ILL and can be accessed via the experiments 9-13-952 [39] and 1-20-69 [40]. The BASIS data were collected during the experiment IPTS 18578.1.

C.B., K.P, I.M., F.R.-R. and T.S. thank the organizers of the QENS/WINS 2022 conference for an excellent meeting in San Sebastian. We thank M. Hennig and M. Grimaldo for earlier versions of the employed MATLAB code. We thank the DFG, ANR (ANR-21-CE06-0047), BMBF (ErUM-pro 05K19VTB and 05K22VTA), and InnovaXN (EU MSCA COFUND 847439) for funding. A portion of this research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory. The authors gratefully acknowledge the financial support provided by JCNS to perform the neutron scattering measurements at the SNS.

References

[1] J. Orear, *Notes on statistics for physicists* (University of California, Lawrence Radiation Laboratory; printed for the US Atomic Energy Commission, 1958)

[2] E.L. Frome, M.H. Kutner, J.J. Beauchamp, Journal of the American Statistical Association **68**, 935 (1973)

[3] S. Baker, R.D. Cousins, Nuclear Instruments and Methods in Physics Research **221**, 437 (1984)

[4] M. Appel, B. Frick, Review of Scientific Instruments **88**, 036105 (2017)

[5] N. Martinez, F. Natali, J. Peters, EPJ Web of Conferences **83**, 03010 (2015)

[6] S. Mukhopadhyay, B. Hewer, S. Howells, A. Markvardsen, Physica B: Condensed Matter **563**, 41 (2019)

[7] E. Mamontov, R.W. Smith, J.J. Billings, A.J. Ramirez-Cuesta, Physica B: Condensed Matter **566**, 50 (2019)

[8] M. Grimaldo, F. Roosen-Runge, F. Zhang, F. Schreiber, T. Seydel, Quarterly Reviews of Biophysics **52**, e7 (2019)

[9] F. Roosen-Runge, D.J. Bicout, J.L. Barrat, Journal of Chemical Physics **144**, 204109 (2016)

[10] G. Kneller, K. Hinsen, The Journal of Chemical Physics **121**, 10278 (2004)

[11] G.R. Kneller, Physical Chemistry Chemical Physics **7**, 2641 (2005)

[12] M. Pathan, M. Garg, S. Mittal, East West Math **8**, 49 (2006)

[13] A.N. Hassani, A.M. Stadler, G.R. Kneller, The Journal of Chemical Physics **157**, 134103 (2022)

[14] D. Gottlieb, C.W. Shu, SIAM review **39**, 644 (1997)

[15] G. Wertheim, Journal of Electron Spectroscopy and Related Phenomena **6**, 239 (1975)

[16] D. Sivia, C. Carlile, W. Howells, S. König, Physica B: Condensed Matter **182**, 341 (1992)

[17] D. Monserrat, A. Vispa, L. Pardo, R. Tolchenov, S. Mukhopadhyay, F. Fernandez-Alonso, Journal of Physics: Conference Series **663**, 012009 (2015)

[18] O. Arnold, J. Bilheux, J. Borreguero, A. Buts, S. Campbell, L. Chapon, M. Doucet, N. Draper, R.F. Leal, M. Gigg et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **764**, 156 (2014)

[19] H. Paalman, C. Pings, Journal of Applied Physics **33**, 2635 (1962)

[20] M.K. Braun, M. Wolf, O. Matsarskaia, S. Da Vela, F. Roosen-Runge, M. Sztucki, R. Roth, F. Zhang, F. Schreiber, The Journal of Physical Chemistry B **121**, 1731 (2017)

[21] M. Grimaldo, F. Roosen-Runge, N. Jalarvo, M. Zamponi, F. Zanini, M. Hennig, F. Zhang, F. Schreiber, T. Seydel, EPJ Web of Conferences **83**, 02005 (2015)

[22] M. Grimaldo, F. Roosen-Runge, M. Hennig, F. Zanini, F. Zhang, N. Jalarvo, M. Zamponi, F. Schreiber, T. Seydel, Physical Chemistry Chem-

ical Physics **17**, 4645 (2015)

[23] A. Girelli, C. Beck, F. Bäuerle, O. Matsarskaia, R. Maier, F. Zhang, B. Wu, C. Lang, O. Czakkel, T. Seydel et al., Molecular Pharmaceutics **18**, 4162 (2021)

[24] C. Beck, M. Grimaldo, M.K. Braun, L. Bühl, O. Matsarskaia, N.H. Jalarvo, F. Zhang, F. Roosen-Runge, F. Schreiber, T. Seydel, Soft Matter **17**, 8506 (2021)

[25] M. Grimaldo, H. Lopez, C. Beck, F. Roosen-Runge, M. Moulin, J.M. Devos, V. Laux, M. Härtlein, S. Da Vela, R. Schweins et al., The Journal of Physical Chemistry Letters **10**, 1709 (2019)

[26] D.B. Anunciado, V.P. Nyugen, G.B. Hurst, M.J. Doktycz, V. Urban, P. Langan, E. Mamontov, H. O'Neill, The Journal of Physical Chemistry Letters **8**, 1899 (2017)

[27] J. Nickels, H. O'Neill, L. Hong, M. Tyagi, G. Ehlers, K.L. Weiss, Q. Zhang, Z. Yi, E. Mamontov, J. Smith et al., Biophysical Journal **103**, 1566 (2012)

[28] C. Neylon, European Biophysics Journal **37**, 531 (2008)

[29] D.I. Svergun, M.H.J. Koch, Reports on Progress in Physics **66**, 1735 (2003)

[30] C. Beck, M. Grimaldo, F. Roosen-Runge, R. Maier, O. Matsarskaia, M. Braun, B. Sohmen, O. Czakkel, R. Schweins, F. Zhang et al., Crystal Growth & Design **19**, 7036 (2019)

[31] C. Beck, M. Grimaldo, H. Lopez, S. da Vela, B. Sohmen, F. Zhang, M. Oettel, J.L. Barrat, F. Roosen-Runge, F. Schreiber et al., The Journal of Physical Chemistry B **126**, 7400 (2022)

[32] A. Cisse, A.L. Schachner-Nedherer, M. Appel, C. Beck, J. Ollivier, G. Leitinger, R. Prassl, K. Kornmueller, J. Peters, The Journal of Physical Chemistry Letters **12**, 12402 (2021)

[33] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright et al., Nature Methods **17**, 261 (2020)

[34] F. Roosen-Runge, M. Hennig, F. Zhang, R.M.J. Jacobs, M. Sztucki, H. Schober, T. Seydel, F. Schreiber, Proceedings of the National Academy of Sciences (USA) **108**, 11815 (2011)

[35] C. Beck, M. Grimaldo, F. Roosen-Runge, M.K. Braun, F. Zhang, F. Schreiber, T. Seydel, The Journal of Physical Chemistry B **122**, 8343 (2018)

[36] B. Hartke, WIREs Computational Molecular Science **1**, 879 (2011)

[37] D.J. Wales, J.P.K. Doye, The Journal of Physical Chemistry A **101**, 5111 (1997)

[38] Z. Li, H.A. Scheraga, Proceedings of the National Academy of Sciences (USA) **84**, 6611 (1987)

[39] L. Colin, M. Appel, C. Beck, C. Grapentin, O. Matsarskaia, K. Pounot, F. Roosen-Runge, F. Schreiber, T. Seydel, *Diffusive dynamics in crowded solutions containing two types of proteins* (2021), https://doi.ill.fr/10.5291/ILL-DATA.9-13-952

[40] F. Roosen-Runge, O. Matsarskaia, T. Seydel, F. Schreiber, C. Beck, *Developing new analysis methods for fixed window scans for soft colloidal suspensions* (2021), https://doi.ill.fr/10.5291/ILL-DATA.1-20-69