Fourier Transform and Distance Distribution Function: New Insights into Data Analysis.

Shiv Kumar Singh
Department of Pure and Applied Physics, Adamawa state University, MUBI,
Adamawa State, NIGERIA

Abstract

We revisit the standard procedure for analyzing the static light intensity curves I(q) measured in scattering from chemical/biological samples as done in the literature. By utilizing some algebraic properties of Fourier integrals we show that; i) the theoretical forward intensity I(0) extracted via the Glatter/Moore algorithm must be consistently higher than all peaks occurring in the input I(q) curve ; ii) the attenuated oscillations of input I(q) data at intermediate q may be a direct consequence of the presence of a bump in the output distance distribution profile P(r) ; iii) if the input intensity is modified by the Gaussian damping factor exp ($-Kq^2$) then extrapolation of the output results down to K = 0 is liable to become numerically unstable ; and iv) due to the exp ($-Kq^2$) artifact the output distance distribution function profile may exhibit a bump having the correct location but wrong width if K is nonzero.

Keywords: Bump, morphology, Fourier integrals, truncation etc.

I. INTRODUCTION

Static scattering of light or X-rays from chemical/biological systems[1-16] is known to provide very valuable information concerning the static structure/morphology of the system. If the scattered intensity I(q) is measured as a function of the transfer wave number q then a Guinier plot[17] estimates

Proc. of the International Conference on Advances in Electronic Devices and Circuits- EDC 2012
Edited by Dr. R. K. Singh.

ISBN: 978-981-07-3161-8

the gyration radius R_g, the Porod slope[18] α tells about the relative roughness/fractality of the target surface, and the Kratky graph[19,20] of q^2 I(q) describes the comparative rigidity of polymer chains. Especially important in this context is the *distance distribution function* P(r)[21] extracted from the measured I(q) curve via Fourier inversion. Such a procedure of data analysis is supposed to be standard and its few relevant features are summarized in Sec II for convenience. Still there are several issues of practical/theoretical nature which require further attention, and the *aim of the present paper is to gain useful insight into these issues.*

Sections III, IV, V and VI below address the following new questions : i) How can the symbol I(0) (which is the radial integral of P(r)) place an important consistency check on the Glatter/Moore[22,23] inversion algorithm ? ii) Can the attenuated oscillations often observed in experimental I(q) curve at medium q be attributed to the presence of a bump in the underlying P(r) profile ? iii) Can an artificial Gaussian damping factor $\exp(-Kq^2)$ in the modified intensity lead to extrapolation uncertainty in the resulting P(r) shape ? iv) Can the $\exp(-Kq^2)$ artifact cause rather large variations in the spread of the predicted P(r) around its bump ?

While answering these questions we shall use some nice properties of Fourier integrals[24]. Also, the graphs/expressions of Ref.[25] will be frequently quoted since that reference serves as a typical prototype. However, our final conclusions mentioned in italics in Secs. III-VI shall be seen to be of quite general validity.

II. STANDARD METHOD OF ANALYSING I(q) DATA

The method essentially proceeds via the following six steps:

1st Step. The experimentalist measures his raw scattered intensities $I_{raw}(q)$ in a photon counter over a truncated region $q_L \leq q \leq q_H$ of the transfer wave number. The termination at the

lower end q_L and higher end q_H is necessitated by the angular limitation of the spectrometer and wavelength restriction of the radiation source.

2nd Step. By employing a multiplicative factor A the raw intensities are renormalized according to

$$I(q) = A\, I_{raw}(q) \qquad ; q_L \le q \le q_H \tag{1}$$

These I(q)'s are now suitable for plotting on a graph and feeding into the computer for further treatment. Obviously the role of A in (1) is only cosmetic. Fig. 1 shows a typical normalized I(q) curve obtained by Hirai et al[25] who employed small angle X-ray scattering to study the thermal conformational change of lysozyme depending on protein concentration and pH. The lower and higher truncation points in this graph are $q_L \sim 0.02$ Å^{-1} and $q_H \sim 0.35$ Å^{-1}.

3rd Step. Suppose in the domain of very small q the I(q) curve drops as q increases. Then a least-square fit of the unknown forward intensity I(0) and gyration radius R_g can be carried out based on the Gunier formula

$$I(q) = I(0)\{1 - q^2 R_g^2/3 + \ \ldots\}\ ; \quad qR_g << 1 \tag{2}$$

Unfortunately, (2) often suffers from many difficulties such as availability of few data points, instability of extrapolation, appearance of nearby peaks in I(q), occurrence of aggregation[26], etc. Indeed, (2) cannot be applied to Fig. 1 which exhibits a small-angle peak at $q_p \sim 0.08$ Å^{-1} on account of a weak intermolecular repulsion between the proteins.

4th Step. In order to reduce the effect of truncation at the higher wave number end q_H, the I(q) curve of (1) is replaced by[25]

$$I(q) \to I(q)\, e^{-Kq^2} \qquad ; \quad K > 0 \tag{3}$$

where K is an artificial damping parameter which is extrapolated to zero at the end of the analysis. Further implications of the recipe (3) will be discussed later in Secs V and V1.

5th Step. One now turns to the question of extracting the distance distribution function $P(r) \equiv 4 \pi r^2 G(r)$ with G(r) being the density-density correlation function[27] of the target system. For this purpose powerful numerical algorithms have been discussed by Glatter[22], Moore[23] and Hansen & Pedersen[28]. Here P(r) is sharply cutoff at distance D i.e. $P(r) = 0$ for $r \geq D$ and it is expanded in a basis set for $0 \leq r \leq D$. Next, the defining Fourier integral

$$I(q) = \int_0^D dr \frac{\sin qr}{qr} . P(r) \tag{4}$$

is set-up at the data points and the unknown expansion coefficients of P(r) are determined via a least-square fit to I(q).

6th Step. Finally, the complete *theoretical* profiles of P(r) and I(q) in the ranges $0 \leq r \leq D$, $0 \leq q \leq \infty$ are plotted leading also to the determination of the quantities

$$I(0) = \int_0^D dr \; P(r) \tag{5}$$

$$R_g^2 = \int_0^D dr \; r^2 \; P(r) / 2 I(0) \tag{6}$$

through numerical quadrature. We are now ready to address the new questions raised in the Introduction.

III. CONSISTENCY CHECK ON I(0)

The *first new insight* is concerned with normalization of the input and output intensities. In sharp contrast to the ad hoc multiplicative factor A of (1) the symbol I(0) of (5) has a deep theoretical significance viz. it represents the normalized scattered intensity at zero angle. Of course, this quantity was experimentally inaccessible due to termination at q_L in Fig. 1 but now it is known from the Glatter quadrature. The integrals (5, 4) obey the inequality

$$I(0) > \int_0^D dr \left| \frac{\sin qr}{qr} \right| P(r) > I(q) \quad ; \quad q > 0 \tag{7}$$

due to the property $1 > |\sin qr|/qr$ for q > 0. We stress that (7) is not to be regarded as a trivial property of Fourier integrals. Its implication is highly nontrivial because I(0) on the left-hand-side is extracted from Glatter's algorithm while I(q) on the right-hand-side is read-off directly from experimental data. In particular, we have $I(0) > I(q_p)$ where $q_p \sim 0.08$ Å^{-1} is the location of the peak in Fig. 1. Stated in words, *(7) tells that the theoretical I(0) of (5) must be consistently higher than all peaks occurring in the input I(q) curve.* Such a check on the numerical procedure was not always adopted in Ref.[25].

IV. OSCILLATIONS OF INTENSITY

The *second new insight* deals with the reason why the I(q) curve in Fig. 1 shows attenuated oscillations, i.e., appearance of crest and trough with steadily decreasing amplitude. For this purpose, we reproduce in Fig. 2 the output P(r) profile obtained in[25] using the Glatter method. Clearly P(r) in Fig. 2 has an almost bell-shaped bump at $r_b \sim 20$ Å which is expected to have a marked influence on the behaviour of I(q) at intermediate q values of order r_b^{-1}. Assuming that the bump at r_b dominates the Fourier integrals (4) we get a crude estimate

$$I(q) \approx I(0)\,\frac{\sin qr_b}{qr_b} + \ldots \quad ; \quad q \sim r_b^{-1} \tag{8}$$

where the dots … stand for some nonleading background terms. Now, the sin qr_b factor in (8) has maxima and minima at

$$\begin{aligned} q &\sim \pi/2\,r_b,\ 3\pi/2r_b,\ \ldots. \\ &\sim 0.08,\ 0.24,\ \ldots\ \text{Å}^{-1} \end{aligned} \tag{9}$$

which almost coincide with the positions of the crest and trough in Fig. 1 if $r_b \approx 20$ Å. Also, the magnitude of the estimate (8) decreases steadily as q increases due to the $1/q\,r_b$ factor. Of course, features in the direct P(r) space and reciprocal I(q) space ought to be mutually related as happens in the case of diffraction patterns produced by crystal lattices/simple liquids whose radial distribution

function G(r) is periodic/quasi periodic. However, the standard literature [26-27] does not report clear formulas like (8,9) applicable to the biological systems whose P(r) may possess a bump at finite r_b. Hence we infer that *the damped oscillations of input I(q) curve at intermediate q may be a direct consequence of the presence of a bump in the output P(r) profile.* This fact has not been emphasized in[25].

V. EXTRAPOLATION IN K PARAMETER

The *third new insight* focuses attention on the modified factor $e^{-K\,q2}$ of (3). One may argue that it is more preferable to employ the pure Glatter method which fits via least squares the directly measured I(q) curve without introducing artificial damping or fictitious data points. However, the pure Glatter method applied over the truncated domain $q_L \le q \le q_H$ sacrifices some accuracy because degrees of freedom are reduced and the unmeasured/excluded data points lie on the possibly rising portion of the I(q) curve in Fig.1. The dual objectives of the Japnese group [25] in using the recipe (3) was to increase the degrees of freedom and also to lessen the effect of the unmeasured portion. For minimizing the effects of truncation at the higher end q_H ~ 0.35 $Å^{-1}$ it is essential to choose

$$K\,q_H^2 > 1 \quad ; \quad K > 10\,Å^2 \tag{10}$$

which implies that numerical analysis based on the modified intensity (3) involves rather large starting K values. Hence *extrapolation of the output results to K = 0 is liable to become inaccurate since the domain 0 ≤ K ≤ 10 $Å^2$ is very wide.* The P(r) profile of Fig. 2 borrowed from[25] is silent about this possible difficulty.

VI. EFFECT OF K-DEPENDENT WIDTH

The *fourth new insight* brings out another peculiarity of the artifact (3) if written as

$$I_M(q) = I(q)\eta(q) \;;\; \eta(q) \equiv e^{-Kq^2} \tag{11}$$

where I_M is the modified intensity, I the actual intensity, and η the artificial damping factor. Clearly, since I_M of (11) is a direct product in the wave number space its inverse Fourier transform must be a convolution integral in the position space (see the Appendix for details). More explicitly, for distances comparable to r_b we have

$$P_M(r) = \int_0^D ds \frac{r}{s} P(s)\xi_C(t) \;;\; t \equiv r - s \tag{12}$$

$$\xi_C(t) \equiv \frac{1}{\pi}\int_0^\infty dq \cos qt.\eta(t) = \frac{e^{-t^2/4K}}{\sqrt{4\pi K}} \tag{13}$$

Here s is an integration variable, $P_M(r)$ is the distance distribution function of the modified intensity, and $\xi_C(t)$ is the cosine Fourier transform of η(q) defined by (11). Although the expression (12) looks complicated yet its behaviour may be easily understood by remembering that P(s) of Fig. 2 has a bump at $s = r_b$ with spread Δ_b (say) and $\xi_C(t)$ of (13) has a peak at t = 0 with width $\sqrt{4K}$. Two limiting cases now become relevant.

<u>*Case (i)*</u>. Suppose K is so small that $\sqrt{4K} << \Delta_b$. Then the function P(s)varies very slowly compared to the factor $\xi_C(t)$ in (12). This yields a desirable result viz.

$$P_M(r) \approx P(r)\int_0^D ds\,\xi_c(t) \approx P(r) \;;\; \sqrt{4K} << \Delta_b \tag{14}$$

<u>*Case (ii)*</u>. Next, let K be so large that $\sqrt{4K} >> \Delta_b$. Then the factor $\xi_C(t)$ varies much slowly compared to the function P(s) in (12). This leads to a peculiar result namely

$$P_M(r) \approx \frac{r}{r_b}\xi_C(r - r_b)\int_0^D ds\, P(s)$$

$$\approx \frac{I(0)}{\sqrt{4\pi K}} \exp\left\{-\frac{(r - r_b)^2}{4K}\right\}; \qquad \sqrt{4K} >> \Delta_b \tag{15}$$

Eqs. (14,15) imply that *if the Glatter/Moore algorithms are applied to the modified $I_M(q)$ of (11) then the resulting $P_M(r)$, though peaked at r_b, will have correct spread if* K is small *but wrong width if K is large*. This feature has not been brought out in Ref.[25].

Towards the end it may be remarked that our review of the standard procedure for data analysis in Sec. II is systematic, properties of the Fourier integrals used in Secs III-VI are sound, physical arguments presented therein are logical, and the Appendix on the convolution formula is neat. The main conclusions of the present work have already been mentioned in the Abstarct.

Conclusion:

The work gives the so many new insight are related to the works. In first new insight is concerned with normalization of the input of the input and output intensities. In particular when I (0) > IC (qp) when qp $0.08A^{-1}$ is the location of the peak in fig2. stated in words (7) tells that the theoretical I(0) of (5) must be consistently higher than all peaks occurring in the input I(9) curve.

The second new insight deals with reason why the I(q) curve in fig2, shows alternated oscillate i.e. appearance of crest and through with steadily decreasing amplitude amenity to the fig (2) thou tout profile obtained in 25 carry the Glatter method clearly in fig (2) has an almost bell shaped bump at B $20A^0$ but in our work damped oscillation of input I(q) curve at intermediate q may be a direct consequence of the presence of a bump in the output $P_{(0)}$ profit.

The third new insight focuses attention on the modified factor e^{-k} q^2 of (3). One may argue that it is more preferable to employ the pure Glatter. Method which fits via least squane the directly measured. I(q) curve without introducing artificial damping and fictitious data points. But have extrapolation of the output result to k = o is liable to become. Inaccurate since the domain $0 \leq k \leq 10A^2$ is very wide. The pits profile of fit (2) borrowed from 25 is silent about this possible difficulty.

The forth and lost work is here that if the Glatter/moore algonithms are applied to the modyl Im(q) of (11) rgen the resulting pimo though peaked at r_b will have corrected. Spread of k is small but wrong width if k is large. This feature has not been brought out in t-r-y-25.

References:

[1] Brown J.C., Pusey P.N., Goodwin J.W. and Ottewill R.H., (1975) Light scattering study of dynamic and time-averaged correlations in dispersions of charged particles J. Phys. Vol.A8, page 664.

[2] Pusey P.N., (1975) The dynamics of interacting Brownian particles. J.Phys. vol. A8,page 1433.

[3] Pusey P.N., (1978) Intensity fluctuation spectroscopy of charged Brownian particles: the coherent scattering function, J. Phys. Vol. A11, page119.

[4]Shaefer D.W., (1977) Quasielastic light scattering by biopolymers. VI. Diffusion of mononucleosomes and oligonucleosomes in the presence of static and sinusoidal electric fields, J. Chem. Phys. Vol 66, page 3980.

[5] Hartl W.and Versmold H., (1982) Structure factor for a two component mixture of dilute colloidal suspensions, J. Chem. Phys. Vol 80, page 1387.

6 Lindsay H.M. and Chaikin P., (1986) Elastic properties of colloidal crystals and glasses, J. Chem. Phys. Vol 76, page 3774.

[7] Pusey P.N. and Megan W.V., (1986) Phase behaviour of concentrated suspensions of nearly hard colloidal spheres, Nature. Vol. 320, page 340.

[8] Pusey P.N. and Megan W.V., (1987) Phase behaviour of concentrated suspensions of nearly hard colloidal spheres, Phys. Rev. Lett. vol 59, page 2083.

[9] Latimer P., (1983) Blood platelet aggregometer: predicted effects of aggregation, photometer geometry, and multiple scattering, Appl. Opt.Vol. 22, page 1136.

[10] Latimer P., (1985) Experimental tests of a theoretical method for predicting light scattering by aggregates, Appl. Opt. Vol. 24, Page 3231.

[11] David P.J., Nair A.C., Menon V.J. and Tripathi D.N., (1996) Laser light scattering studies from blood platelets and their aggregates , Coll. Surf. B: Biointerfaces vol. 6, page 101.

[12] Harold D.B. and Schmidt P.W., (1984) New insights into optical Fourier transforms, radial distribution function, and Glatter's method Phys. Rev. Lett vol.53,page 596.

[13] Hirai M., Niimura N., Zama M., Mita K., Ichimura S., Tokunaga F. and Ishikawa Y., (1988) Interparticle interactions and structural changes of nucleosome core particles in low-salt solution, Biochemistry. Vol. 27,Page 7924.

[14] Dubey Ritesh Kumar, Pandey R.K. and Tripathi D.N., (2002) Proton/antiproton collisions with positronium in the presence,Ind. J. Pure & Appl. Phys.Vol 40,Page 111.

[15] Dubey Ritesh Kumar and Tripathi D.N., (2004) Ordering in Polystyrene Macroions, Physica Scripta Vol 69, Page 461.

[16] Schaefer D.W. and Keefer K.D., (1986) Small-angle X-ray scattering study of cordierite sol-gel synthesis Phys. Rev. Lett. Vol 56,Page 2199.

[17] Guinier A., (1939), Atomic structure of pre-Guinier-Preston zones in Al alloys, Ann. Phys. Vol 12,Page 161.

[18] Porod G., (1951), Small-angle x-ray scattering of high- and low-affinity heparin,Kolloid Z. Vol. 124,Page 251.

[19] Kratky O., Porod G., (1949) The worm-like chain (WLC) model in polymer physics is used to describe the behavior of semi-flexible polymers, Recl. Trav. Chim. Pays-Bas vol. 68, Page 1106.

[20] Kriste R.G., Oberthur R.C., O. Glatter, O. Kratky (1982) In Small Angle X-ray Scattering, Eds. Academic Press London, p. 387.

[21] The distance distribution function P(r) for a molecular system of number density n is defined as follows. If around an arbitrary molecule 1 we consider a spherical shell of radius r and thickness dr then n P(r)dr gives the average number of molecules inside this shell. The P(r) function depends on the particle shape as well as their arrangement.

[22] Glatter O., (1977) A new method for the evaluation of small-angle scattering data, J. Appl. Cryst. Vol. 10, Page 415.

[23] Moore P.B., (1980), Small-angle scattering. Information content and error analysis, J. Appl. Cryst. Vol. 13,Page 168.

[24] Titchmarsh E.C.,(1939) The Theory of Functions, Second Edn, Oxford Univ. Press, Oxford, Chap XIII.,

[25] Hirai M., Arai S., Iwase H. and Takizawa T., (1998) , Small-angle X-ray scattering and calorimetric studies of thermal conformational change of lysozyme depending on pH, J. Phys. Chem B Vol.102,.

[26] Feigin L.A., Svergun D.I., Taylor Ed. G.W. (1987) In Structure Analysis by Small-Angle X-ray and Neutron Scattering, (Plennum Press New York, Chapt 3, Chapt 9,).

[27] Kerker M., (1969) The Scattering of Light and Other Electromagnetic Radiation, (Academic Press, New York and London, p. 451, 459-461,).

[28] Hansen S., Pedersen J.S., (1991) Small-angle X-ray scattering of systems with inhomogeneous particles, J. Appl. Cryst. Vol.24, Page 541.

Figure Captions:

Figure 1. Prototype experimental X-ray scattering curves I(q) of the hen egg-white lysozyme in hepes buffer at various temperatures for pH 7.0 and protein concentration 5% w/v as measured by Hirai et al[25].

Figure 2. Distance distribution function P(r) obtained by Hirai et al[25] using Glatter algorithm depending on temperature for pH 7.0 and protein concentration 5%w/v.

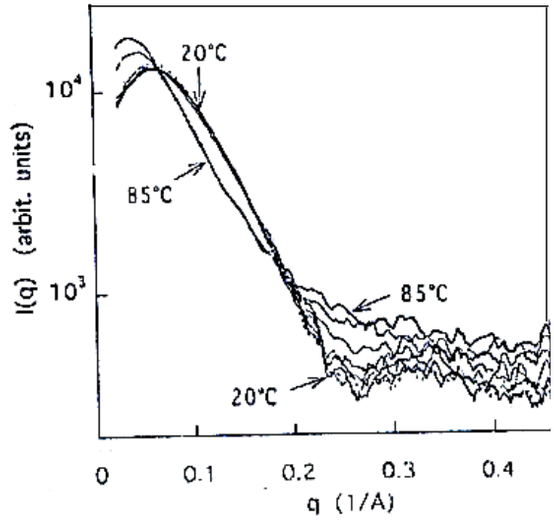

Fig. 1

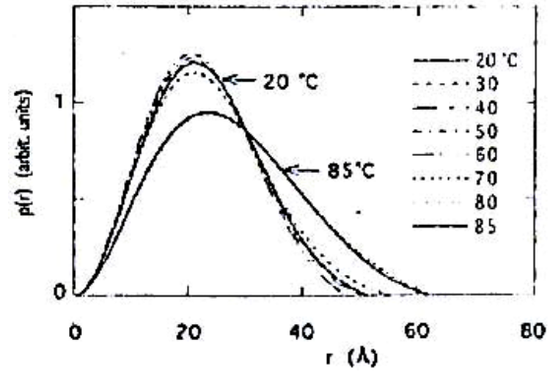

Fig. 2