Abstract

The aim of this paper is to assess the potential accuracy of structure analyses based upon solution X-ray scattering experiments, before the structure problem is solved. In other words, an algorithm is sought expressing accuracy in terms of the experimental observations and associated errors, independently of the value of the structure parameters. The mathematical treatment involves five steps: (a) Data reduction, namely the definition in experimental space of the degrees of freedom of the problem and the determination of their probability law. (b) Choice of a representation of the structure consistent with the coarse resolution of solution scattering experiments; a set of sharp pseudo-atoms is adopted, so that the degrees of freedom in real space are the position and electron density of each pseudo-atom. (c) Introduction of a 'virtual experiment' defined by a noise distorting the structure parameters, and determination of the corresponding probability law in the experimental space. (d) Comparison of the real and the virtual experiments, or more precisely of the probability laws of the degrees of freedom in experimental space, yielding a relation between the variances of the position and electron density uncertainties, the number of the pseudo-atoms and the value of the structure parameters. (e) Attenuation of the structure dependence via the introduction of a hypothesis of quasi-spherical symmetry, whose effect is to reduce the expression of the variances to a function of the experimental observations, independent of the knowledge of the structure of the particle. This analysis is applied to solution X-ray scattering experiments at variable contrast, interpreted within the framework of the invariant volume hypothesis. The final result takes the form of a curve relating the variances of the position and electron density uncertainties, divided by the number of pseudo-atoms. The example of a low-density serum lipoprotein is used to illustrate these results.

I. Introduction

In this series of papers we address the problem of estimating accuracy and resolution in structural studies based upon scattering techniques. Ideally, our purpose is to elaborate mathematical criteria involving only the experimental observations - i.e. not requiring the prior knowledge of the structure under investigation - applicable to solution scattering as well as to crystallographic studies, and thus to set objective standards to assess the intrinsic quality of any particular structural analysis at any stage of progress. In other words, we seek explicit correlations between the probability law of the experimental observations and that of the structure parameters. Although our analysis is focused on X-ray scattering studies, the results can easily be extended to other radiations (particularly neutrons).

The first two papers of this series (Luzzati, 1980; Taupin & Luzzati, 1982) are devoted to the problem of data reduction in solution scattering studies; the third one (Luzzati & Taupin, 1984) deals with the assessment of accuracy and resolution in crystallographic structural analyses. This and the accompanying paper are focused on the problem of accuracy in low-resolution 'small-angle' scattering studies: this paper deals with solution scattering, the accompanying one with crystallographic structure analyses restricted to the same 'small-angle' range as the solution studies.

Essential for our analysis is the notion that the separation of reciprocal space into a 'small-angle' and a 'high-angle' region makes clear physical sense. This separation, amply documented in solution scattering literature, is justified by the very existence in real space of two distinct regimes of scattering density fluctuations: one - typically in the range 1 to 10 Å - corresponds to the short interatomic distances (chemical bonds, van der Waals contacts etc.), the other to the structural features at the macromolecular level (see Luzzati & Tardieu, 1980). The separation between the
two regions is often and empirically set at approximately \((25 \text{ Å})^{-1}\) and Porod's law is used to bridge the gap.

The analysis developed in this and in the accompanying paper can be decomposed into five logical steps:

(a) **Data reduction**, namely the definition in reciprocal space of the degrees of freedom of the problem and the determination of their probability laws.

(b) **Representation of the structure**, namely the choice in real space of a parametrization of the structure consistent with the coarse resolution of the small-angle scattering phenomenon.

(c) **Introduction of a virtual experiment**, defined in real space by a stochastic distortion of the structural parameters (a 'noise'), and subsequently determination in reciprocal space of the corresponding probability law of the degrees of freedom.

(d) **Comparison of the real and the virtual experiments**, or more precisely of the probability laws of the degrees of freedom in reciprocal space. This comparison sustains the determination of some stochastic parameters of the 'noise' and thus yields a statistical definition of accuracy and resolution.

(e) **Attenuation of the structure dependence**, namely introduction of a few hypotheses that lead to an expression of accuracy and resolution involving only experimental observations and not the knowledge of the structure of the particle.

The physical and the mathematical problems involved in each of those five steps are treated in §§ II.1 to II.6. A less technical analysis of those problems is presented in § IV.

**II. Mathematical treatment**

Most of the mathematical background is presented in the previous papers of this series (Taupin & Luzzati, 1982; Luzzati & Taupin, 1984). For the sake of clarity we recall here the notation and some of the basic notions involved in this and in the accompanying paper.

**II.1. Notation**

\[ M \text{ relational operator meaning 'is the } M\text{-dimensional Fourier transform of'} \]

\[ \ast \text{ convolution operator: } (f \ast g)(x) = \int f(x-y)g(y) \, dy. \]

\[ \langle f \rangle \text{ spherical average of some function } f \text{ of a three-dimensional space. } \langle f(r) \rangle \text{ is a function of the scalar } r = |r|. \]

\[ r, r \text{ a position in real space, } r = \{x, y, z\}, \quad r = |r|. \]

\[ \text{dr, dX etc. } \text{d denotes a differential volume element in some vectorial space } r, X \text{ etc.} \]

\[ s, s \text{ a position in the reciprocal space, } s = |s|, s = (2 \sin \theta)/\lambda. \]

\[ h, h \text{ labels in reciprocal space: } h \text{ identifies reciprocal-lattice points in crystallographic studies, } h \text{ identifies Shannon's channels in solution scattering studies.} \]

\[ N, n \text{ number of atoms and of pseudo-atoms per (hydrated) particle.} \]

\[ \hat{\rho} \text{ solvent electron density.} \]

\[ \rho_1 \text{ solvent electron density at particle buoyancy.} \]

\[ v_1(r), v_1 \text{ (hydrated) particle shape function and particle volume: } v_1 = \int v_1(r) \, dr. \text{ We suppose here that the function } v_1(r) \text{ takes one of two values, 0 or 1 (Luzzati & Tardieu, 1980).} \]

\[ m \text{ number of electrons of one (anhydrous) particle.} \]

\[ MW \text{ weight (in daltons) of one (anhydrous) particle.} \]

\[ \Delta \rho_1(r, \hat{\rho}) \text{ particle electron density contrast at solvent density } \hat{\rho}. \]

\[ \Delta \rho_1(r) \text{ particle electron density contrast at solvent electron density } \rho_1 \text{ (namely at buoyancy); the invariant volume hypothesis states that (Luzzati & Tardieu, 1980)} \]

\[ \Delta \rho_1(r, \hat{\rho}) = \Delta \rho_1(r) - (\hat{\rho} - \rho_1)v_1(r). \quad (\text{II-1-1}) \]

\[ M_{\mu,v} \mu \text{th moment, with respect to the centre of gravity of the particle, of the } v \text{th power of the electron density contrast between particle and solvent, at buoyancy (by definition of buoyancy } M_{0,1} = 0): \]

\[ M_{\mu,v} = \int r^\mu \Delta \rho_1^v(r) \, dr \]

\[ = 4\pi \int_0^{D_{\text{av}}/2} r^\mu + 2\langle \Delta \rho_1^v(r) \rangle \, dr. \quad (\text{II-1-2}) \]

\[ p(r, \hat{\rho}) \text{ particle one-dimensional autocorrelation function at solvent density } \hat{\rho}: \]

\[ p(r, \hat{\rho}) = \langle \Delta \rho_1(r, \hat{\rho}) \ast \Delta \rho_1(-r, \hat{\rho}) \rangle. \quad (\text{II-1-3}) \]

\[ s \text{ intensity corresponding to } p(r, \hat{\rho}): \]

\[ s(s, \hat{\rho}) [\text{FT}] r p(r, \hat{\rho}). \quad (\text{II-1-4}) \]

\[ \xi(s) \text{ Stuhrmann’s characteristic functions: } \xi \in \{\rho, \rho v, v\}. \]

\[ \text{Patterson function corresponding to } i_\xi(s) \text{ [see (II-1-3) and (II-1-4)]:} \]

\[ p_\xi(r) = \langle \Delta \rho_1(r) \ast \Delta \rho_1(-r) \rangle \quad (\text{II-1-5a}) \]
\[ p_\mu(r) = \langle [\Delta \rho_1(r) \ast v_i(-r)] + [\Delta \rho_1(-r) \ast v_i(r)] \rangle \]
\[ p_i(r) = \langle v_i(r) \ast v_i(-r) \rangle. \]  

(II-1-5b)

\[ s_m \]
value of s at which we set the separation between the 'small-angle' region and the rest of reciprocal space (note that \( s_m \) has a different meaning in Luzzati & Taupin, 1984).

\( X \)
a vector representing (with various sub- and superscripts) in reciprocal space the parameters of the problem, namely the (finite) set of mathematical entities necessary and sufficient for an exhaustive description of the problem.

\( Y \)
a vector representing (with various sub- and superscripts) all the possible values of the experimental observation.

\( \rho(X) \)
a priori probability density of \( X \), before any experiment.

\( q(Y|X) \)
probability law of \( Y \) knowing \( X \).

\( r(X|Y) \)
probability density of \( X \) knowing \( Y \); its expression in terms of \( t(X) \) and \( q(Y|X) \) is
\[ r(X|Y) = q(Y|X)t(X)/\int q(Y|X)t(X) \, dX. \]  

(II-1-6)

\( H \)
(Kullback’s) information corresponding to the event that the probability density of \( X \), which a priori was \( t(X) \), becomes \( r(X|Y) \) after the observation \( Y \):
\[ H = \int r(X|Y) \log_2 \left[ r(X|Y)/t(X) \right] \, dX. \]  

(II-1-7)

\( \mathcal{N}(X, \Sigma) \)
the normal (Gaussian) probability law in an \( M \)-dimensional space:
\[ \mathcal{N}(X, \Sigma) = [(2\pi)^M |\Sigma|]^{-1/2} \exp[-\frac{1}{2} X \Sigma^{-1} X]. \]  

(II-1-8)

\( \mathcal{E}(W|R) \)
the mathematical expectation of some function \( W(x) \) under some probability density \( R(x) \):
\[ \mathcal{E}(W|R) = \int W(x)R(x) \, dx. \]  

(II-1-9)

II.2. Data reduction

The phenomenon we are concerned with is the scattering of an X-ray beam by a solution of macromolecules in a solvent of variable density. Each subexperiment (subexperiments are characterized by distinct beam intensities, concentrations etc. and are referenced by a label \( \mu \)) corresponds to a subset of the components of the global observed vector \( Y_\circ \) consisting of the counts \( Y_{\circ,k,\mu} \) recorded at the \( k \)th channel and of the scaling factor \( K_{\circ,\mu} \) (see Taupin & Luzzati, 1982):
\[ Y_\circ \equiv \{ Y_{\circ,k,\mu}, K_{\circ,\mu} \}. \]  

(II-2-1)

The very existence of a finite number of degrees of freedom hinges upon the validity of the following conditions (Luzzati, 1980):

(a) the autocorrelation function has a finite support of measure \( D_m \);

(b) beyond some threshold \( s_m \) the ideal intensity curve can be described by an asymptotic form, which is a function of a small number of parameters.

In this case each scattering curve is entirely defined by its values at the one-dimensional lattice:
\[ s_h = h/2D_m; \quad h = 1, 2, \ldots, h_m \]  

(II-2-2a)

\[ h_m = \text{round} \ 2D_m s_m \]  

(II-2-2b)

plus the parameters specifying the asymptotic form.

Moreover, and for the sake of simplicity, it is convenient to introduce the following conditions (which are not strict mathematical requirements):

(c) the sample is an ideal solution of discrete particles;

(d) the particles are globular;

(e) all the particles are identical;

(f) the electron density contrast associated with one particle of solute is linearly dependent upon the electron density of the solvent [invariant volume hypothesis, see (II-1-1)].

If these conditions are all fulfilled, then the intensity \( i(s, \hat{\rho}) \) scattered at the angle \( \hat{\rho} \) in the experiment performed in a solvent of electron density \( \hat{\rho} \) takes the form [see (II-1-4) and (II-1-5) and also Luzzati & Tardieu (1980)]
\[ i(s, \hat{\rho})/K = i_\rho(s)/m + (\hat{\rho} - \rho) i_\rho(s)/m \]
\[ + (\hat{\rho} - \rho)^2 i_\rho(s)/m, \]  

(II-2-3)

where \( K \) is a normalization factor.

Finally, taking into account the two parametrizations [(II-2-2) and (II-2-3)], assuming that the asymptotic form of the intensity curves agrees with Porod's law and disregarding the normalization factors \( K_{\circ,\mu} \), we find that the elements of the vector \( X \) are
\[ X \equiv \{ X_{\circ,h} \}; \quad X_{\circ,h} = i_\circ(h/2D_m)/m; \quad 1 \leq h \leq 2s_mD_m. \]  

(II-2-4)

The data reduction process leads to the determination of the position
\[ X_\circ \equiv \{ X_{\circ,h} \}; \quad X_{\circ,h} = i_\circ(h/2D_m)/m \]  

(II-2-5)

of the maximum of \( q(Y_\circ|X) \) and of the local form of \( q(Y_\circ|X) \) near \( X_\circ \).
II.3. **Representation of the structure**

The electron density distribution corresponding to the 'small-angle' scattering experiments is smooth at the scale of the atomic dimensions: it clearly would be unrealistic to represent this function in terms of the individual atomic positions, as is customary in high-resolution crystallographic analyses. A more convenient representation involves the values of the electron density at a number \( n \) of points \( \{r_j\} \) (the sampling is supposed to be dense and regular, as discussed below). More precisely, we adopt in these two papers a representation consisting of \( n \) sharp pseudo-atoms (delta functions) of position \( r_j \) and content \( \varphi [\rho_j - (\hat{\rho} - \rho_1)] \), where \( \varphi = v_1/n \) is the average volume per pseudo-atom and \( \rho_j \) is the electron density at buoyancy (namely when \( \hat{\rho} = \rho_j \)). \( n \) is assumed to be large \( (n \gg 1) \), but smaller than (or equal to) the number of atoms \( N \). The expression of the intensity corresponding to this representation is

\[
i(s, \hat{\rho}) = \varphi^2 \sum_{j=1}^{n} \sum_{l=1}^{n} [\rho_j - (\hat{\rho} - \rho_1)] \int T_{jl}(s),
\]

where

\[
T_{jl}(s) = \frac{(\sin 2\pi r_{jl}s)}{2\pi r_{jl}s} \quad (II-3-1b)
\]

\[
r_{jl} = |r_j - r_l|.
\]

The expressions for the three characteristic functions are [see (II-1-5)]

\[
i_p(s) = \varphi^2 \sum_{j=1}^{n} \sum_{l=1}^{n} \rho_j \rho_l T_{jl}(s) \quad (II-3-2a)
\]

\[
i_{pe}(s) = -2\varphi^2 \sum_{j=1}^{n} \sum_{l=1}^{n} \rho_j T_{jl}(s) \quad (II-3-2b)
\]

\[
i_s(s) = \varphi^2 \sum_{j=1}^{n} \sum_{l=1}^{n} T_{jl}(s). \quad (II-3-2c)
\]

A consequence of the delta functions is the presence of a constant asymptotic term in the intensity curves [see (II-3-1)]:

\[
\lim_{s \rightarrow \infty} i(s, \hat{\rho}) = \varphi^2 \sum_{j=1}^{n} [\rho_j - (\hat{\rho} - \rho_1)]^2, \quad (II-3-3)
\]

which is at variance with Porod's law. We suppose that \( n \) is sufficiently large so that this constant term is negligible over the range \( s \leq s_m \).

By definition, the sampling is considered dense and regular when (see Luzzati & Tardieu, 1980)

\[
i(0, \hat{\rho}) = \{ \int [\Delta \rho_j(r, \hat{\rho}) - \hat{\rho}] \, dr \}^2
\]

\[\simeq \{ \varphi \sum_{j=1}^{n} [\rho_j - (\hat{\rho} - \rho_1)] \}^2 \quad (II-3-4)\]

\[
p(0, \hat{\rho}) = \int [\Delta \rho_j(r, \hat{\rho}) - \hat{\rho}]^2 \, dr
\]

\[\simeq \varphi \sum_{j=1}^{n} [\rho_j - (\hat{\rho} - \rho_1)]^2. \quad (II-3-5)\]

For the sake of compactness, it is convenient to collect the various real-space variables into a unique vectorial space:

\[
Z \equiv \{\{\rho_j\}, \{x_j\}, \{y_j\}, \{z_j\}\} \equiv \{\{\rho_j\}, \{r_j\}\} \quad (II-3-6)
\]

and to collect (II-2-4) and (II-3-2) into a single function:

\[
X = \Phi(Z). \quad (II-3-7)
\]

II.4. **Virtual experiment**

The virtual experiment consists in virtually observing positions and contents \( Z^* \) of the pseudo-atoms offset from their 'correct' values \( Z \) by a 'noise' \( \Delta Z \). We adopt for these offset virtual positions and contents the following notation:

\[
Z^* \equiv \{\{\rho_j^*\}, \{x_j^*\}, \{y_j^*\}, \{z_j^*\}\}
\]

\[\equiv \{\{\rho_j^*\}, \{r_j^*\}\} \quad (II-4-1a)\]

\[
\Delta Z \equiv \{\{\Delta \rho_j\}, \{\Delta x_j\}, \{\Delta y_j\}, \{\Delta z_j\}\}
\]

\[\equiv \{\{\Delta \rho_j\}, \{\Delta r_j\}\} \quad (II-4-1b)\]

\[
Z^* \equiv Z + \Delta Z. \quad (II-4-1c)
\]

According to our conventions (see § II.1), \( \check{i}_p(h) \), \( \check{i}_{pe}(h) \) and \( \check{i}_s(h) \) corresponding to \( Z^* \) [see (II-3-2)] are the elements of the virtual observation vector \( Y^* \):

\[
Y^* \equiv X^* = \Phi(Z^*). \quad (II-4-2)
\]

The probability density \( P_Z = P_Z(Z^*) \) of \( Z^* \), which depends on \( Z \) and on some other parameters \( U \), can be chosen so that each component of \( \Delta Z \) is an independent stochastic variable of zero expectation:

\[
\delta(\Delta Z|P_Z) = 0 \quad (II-4-3a)
\]

\[
\delta(\Delta Z_j|\Delta Z_k|P_Z) = 0, \quad j \neq k. \quad (II-4-3b)
\]

In addition, we decide that each pseudo-atom has the same isotropic statistical properties and we define the two components \( u \) and \( w \) of the vector \( U \) in terms of the variances \( \sigma^2_{\text{pos}} \) and \( \sigma^2_{\text{chr}} \) of the positional and chromatic parameters:

\[
\delta(\Delta x_j^2|P_Z) = \delta(\Delta y_j^2|P_Z) = \delta(\Delta z_j^2|P_Z)
\]

\[= \frac{1}{N} \delta(|\Delta r_j|^2|P_Z) = \frac{1}{N} \delta(\Delta \rho_j^2|P_Z) = \frac{1}{N} \mu^2 = \frac{1}{N} \sigma^2_{\text{pos}} \quad (II-4-3c)\]

\[
\delta(\Delta \rho_j^2|P_Z) = \mu^2 = \sigma^2_{\text{chr}}. \quad (II-4-3d)\]

We assume that it is possible to approximate the virtual transfer function equivalent to \( P_Z(Z^*) \) in the \( X^* \) space by a normal law.*

*The function \( i(s) \) can be expressed as the interpolated spherical average of the intensity scattered by an ideal crystal. In this case Shannon's interpolation algorithm can be used and \( \check{i}(s) \) expressed as a linear combination of the ideal intensities \( \{l(h)\} \) corresponding to each reciprocal-lattice point; if the deviations of the \( \{l(h)\} \) induced by the offsetting of the pseudo-atom positions can be assumed to be independent of each other then the requirements of the central limit theorem are fulfilled.*
\[ P_X(X^t) = q^t(Y^t|Xc, U) \equiv q^t(X^t|Xc, U) \]
\[ = \mathcal{N}(X^t - Xc, \Sigma^t) \quad (II-4-4) \]
whose parameters \( Xc \) and \( \Sigma^t \) can be expressed using the mathematical properties of the normal law [see (II-1-8) and (II-1-9)]:
\[ \Sigma^t_{ij} = \mathcal{E}((X^t_i - Xc_i)(X^t_j - Xc_j)|P_X(X^t)) \quad (II-4-5a) \]
\[ Xc = \mathcal{E}[P_X(X^t)] \quad (II-4-5b) \]
The above identities (II-4-5) may be expressed in terms of \( Z^t \) using (II-4-2):
\[ P_X(X^t) \, dX^t = P_Z(Z^t) \, dZ^t \quad (II-4-6) \]
so that
\[ \Sigma^t_{ij} = \mathcal{E}([F(Z^t) - Xc][F(Z^t) - Xc], P_Z(Z^t)) \quad (II-4-7a) \]
\[ Xc = \mathcal{E}[P_Z(Z^t)] \quad (II-4-7b) \]
Besides, it must be noted that the centre \( Xc \) of the normal law \( P_X(X^t) \) is not generally equal to \( X \) since
\[ \mathcal{E}[P_Z(Z^t)] = \mathcal{E}[P_Z(Z^t)] + \Delta c \quad (II-4-8a) \]
\[ Xc = X + \Delta c. \quad (II-4-8b) \]
As a rule, \( \Sigma^t \) depends not only on \( U \) and \( Xc \) but also directly on \( Z \), and \( \Delta c \) is a function of \( X \) and of \( Z \) [see (II-4-7a) and (II-4-8)]. Therefore, the virtual transfer function takes the form
\[ q^t(X^t|X, U, Z) = \mathcal{N}(X^t - X - \Delta c(U, X, Z), \Sigma^t(U, X, Z)). \quad (II-4-9) \]

II.5. Bringing together the real and the virtual experiments

\( q(Y|X) \) is an empirical function defined by the experimental observations and the associated variances, \( q^t(X^t|X, U, Z) \) is a function of \( X, Z \) and of the stochastic parameters \( U \). Our purpose is to bring the two functions together and thus gain some information regarding \( U \). The problem of comparing two probability laws is not uniquely defined and can be approached in different ways. We have advocated (Luzzati & Taupin, 1984) a procedure based upon the information \( H \) [see (II-1-7)]: the two probability laws are considered to be equivalent when the information \( H \) and \( H^t(U, X^t, Z) \) associated respectively with the pairs \( q(Y|X) \) and \( q^t(X^t|X, U, Z) \) is equal.

This procedure involves the prior probability density \( t(X) \) – namely the reciprocal-space counterpart of the real-space stochastic model expressing a total ignorance – in addition to \( q(Y|X) \) and \( q^t(X^t|X, U, Z) \). In fact, the determination of a satisfactory expression of \( t(X) \) is by no means a trivial problem, since the validity of the expansions most commonly used is restricted to the vicinity of their centre (Bricogne, 1984); more precisely, the Gaussian approximations of both \( t(X) \) and \( q(Y|X) \) or both \( t(X) \) and \( q^t(X^t|X, U, Z) \) cannot be used at the same time unless the centres of the three functions are close to each other. One way to by-pass this problem is to introduce the less-stringent hypothesis that \( t(X) \) is much flatter than both \( q(Y|X) \) and \( q^t(X^t|X, U, Z) \). In this case, the functions \( q(Y|X) \) and \( r(X|Y) \) coincide, as well as the functions \( q^t(X^t|X, U, Z) \) and \( r^t(X^t|X, U, Z) \) [save for a scaling factor, see (II-1-6)]. Therefore, the function \( r(X|Y) \) takes the form
\[ r(X|Y) = \mathcal{N}(X - Xo, \Sigma) \quad (II-5-1) \]
so that the expression for the information \( H \) becomes [see (II-1-7)]
\[ H = -\log_2[t(Xo)] - \frac{1}{2} \log_2|\Sigma| - (M/2)\log_2(2\pi e), \quad (II-5-2) \]
where \( M \) is the number of dimensions of the \( X \) space. In general, the expression for \( r^t(X|X^t, U, Z) \) [see (II-4-9)]
\[ r^t(X|X^t, U, Z) = q^t(X^t|X, U, Z)/q^t(X^t|X, U, Z) \, dX \]
(II-5-3) is not a normal law of the variable \( X \), since \( \Delta c \) and \( \Sigma^t \) are also functions of \( X \). This problem can be circumvented by assuming, as commonly done in statistical analyses [see equation (II-1-4b) in Taupin & Luzzati (1982)], that the values of \( \Delta c \) and \( \Sigma^t \) are almost constant within the range of \( X \) over which the normal law (II-4-9) takes non-negligible values:
\[ \Delta c(U, X, Z) = \Delta c(U, X^t, Z) \quad (II-5-4a) \]
\[ \Sigma^t(U, X, Z) = \Sigma^t(U, X^t, Z) \quad (II-5-4b) \]
In this case, the approximate expressions for \( r^t(X|X^t, U, Z) \) and for \( H^t(U, X^t, Z) \) take the form
\[ r^t(X|X^t, U, Z) = \mathcal{N}([X - Xo] + \Delta c(U, X^t, Z), \Sigma^t(U, X^t, Z)) \]
(II-5-5a)
\[ H^t(U, X^t, Z) = -\log_2\{t(X^t - \Delta c(U, X^t, Z))\} \]
\[ - \frac{1}{2} \log_2|\Sigma^t(U, X^t, Z)| - (M/2)\log_2(2\pi e). \quad (II-5-5b) \]

Bringing together the real and the virtual experiment will then consist in stating that the virtually observed vector \( X^t \) is equal to the maximum \( Xo \) of the transfer function \( q(Y|X) \) and that the two experiments convey the same information. Since, moreover, \( t(X^t) \approx r(X^t - \Delta c) \) [this is a consequence of \( t(X) \) being flat], then the result of equalling the information in (II-5-2) and (II-5-5b) is
\[ |\Sigma^t(U, X^t, Z)| = |\Sigma|. \quad (II-5-6) \]
It must be stressed that (II-5-2) and (II-5-5b) apply to the \(M\)-dimensional space of the vector \(X\) and that the number \(M\) is specified by the nature of the problem (see § II.2). The application of the statistical analysis to a subspace of \(X\) must be justified by a proper reformulation of the probability laws \(q(Y|X)\) and \(t(X)\). More specifically, and from a more practical viewpoint, the statistical analysis can be applied to a single intensity curve, corresponding to a particular solvent density \(\bar{\rho}_b\). This operation is equivalent to foregoing the invariant volume hypothesis and to relying upon an equation similar to (II-2-3):

\[
ii(h, \bar{\rho})/K = \ii(h, \bar{\rho}_b)/m + (\bar{\rho} - \bar{\rho}_b) \frac{f_1(h)}{m} + (\bar{\rho} - \bar{\rho}_b)^2 \frac{f_2(h)}{m}, \text{(II-5-7)}
\]

in which only the variable \(\ii(h, \bar{\rho}_b)\) is a function of structural parameters [see (II-3-1a)], whereas the terms \(f_1(h)\) and \(f_2(h)\) play the role of interpolation parameters unrelated to the structure. These considerations apply to the intensity curve corresponding to any solvent density, for example to \(\ii(h)(\bar{\rho}_b----\bar{\rho}_c)\).

Besides, it is sometimes convenient in error analysis to sacrifice mathematical rigour and to treat each degree of freedom – in the circumstance the value of one characteristic function at one channel \(h\) – independently of all the others, i.e. to neglect cross correlations in the offset probability laws. It is possible to adapt the mathematical procedure above to this situation by adopting a new convention: \(q(Y|X)\) and \(q(X|U, Z)\) are considered to be equivalent when the products of the diagonal terms of the \(S\) matrices are equal:

\[
\prod_{i=1}^{M} \sigma_i^2 = \prod_{i=1}^{M} \delta[(X_i - X_c)^2|P_X]. \text{(II-5-8)}
\]

It is worth noting that if the cross terms are more important in the matrix \(\Sigma\) than in the matrix \(\Sigma^*\), then restricting the analysis to the diagonal terms is likely to entail a deterioration of the parameters \(u\) and \(w\); indeed, neglecting cross terms is a pessimistic alteration of the probability laws. The effects of these approximations are illustrated in § III.

II.6. Expression for the \(Z\) dependence in terms of the experimental parameters \(X_0\)

Equation (II-5-6) is a function not only of \(u\) and \(w\) but also of the structural parameters \(Z\); as a rule, therefore, the values of \(u\) and \(w\) cannot be determined unless the structure of the particle is known. In fact, we are striving here at assessing accuracy and resolution before the structural problem is solved. This difficulty can be circumvented as follows.

Firstly, we introduce the following notation: the structure factor \(f(s)\) of a particle whose electron density distribution is \(g(r)\) can be decomposed into its spherical average \(\sigma(s)\) and its departure from spherical symmetry \(\delta(s)\):

\[
f(s) = \sigma(s) + \delta(s) [FT]^3 g(r) \text{ (II-6-1a)}
\]

\[
\sigma(s) [FT]^3 \langle g(r) \rangle \text{ (II-6-1b)}
\]

\[
\delta(s) [FT]^3 \left[ g(r) - \langle g(r) \rangle \right] \text{ (II-6-1c)}
\]

\[
\delta(s) = \langle \delta(s) \rangle. \text{ (II-6-1d)}
\]

The result of applying (II-6-1) to Stuhrmann's characteristic functions is (see Luzzati, Tardieu, Mateu & Stuhrmann, 1976)

\[
i_{\rho}(s) = \sigma_{\rho}^2(s) + \delta_{\rho}^2(s) \text{ (II-6-2a)}
\]

\[
i_{\rho}(s) = -2[\sigma_{\rho}(s)\sigma_s(s) + \langle \delta_{\rho}(s)\delta_s(s) \rangle] \text{ (II-6-2b)}
\]

\[
i_{\rho}(s) = \sigma_s^2(s) + \delta_s^2(s) \text{ (II-6-2c)}
\]

where the suffixes \(\rho\) and \(v\) refer to the function \(g(r)\) [see (II-6-1)] being equal respectively to \(\Delta \rho_1(r)\) and to \(\nu_1(r)\).

Secondly, we assume that the approximate equalities

\[
\langle (\sigma_{\rho}(s) + \delta_{\rho}(s)) \exp(2\pi ir_1s) \rangle \approx \langle \sigma_{\rho}(s) + \delta_{\rho}(s) \rangle \text{ (II-6-3a)}
\]

\[
\langle (\sigma_s(s) + \delta_s(s)) \exp(2\pi ir_1s) \rangle \approx \langle \sigma_s(s) + \delta_s(s) \rangle \text{ (II-6-3b)}
\]

\[
\langle \delta_{\rho}(s)\delta_s(s) \rangle \approx \delta_{\rho}(s)\delta_s(s) \text{ (II-6-4)}
\]

hold true and that the overall contributions of the products \(\delta_{\rho}(h)\sigma_{\rho}(h), \delta_s(h)\sigma_s(h), \delta_{\rho}(h)\sigma_s(h)\) and \(\delta_s(h)\sigma_{\rho}(h)\) to the determinant \(|\Sigma^*|\) are negligible (see Appendix 1).

These conditions, which are equivalent to a hypothesis of quasi-spherical symmetry, have the effect of transforming the expression of \(|\Sigma^*|\) into a function [see equations (A-1-8) in Appendix 1] of the experimental parameters \(\{X_0, a\}\) and of four sets of structural parameters \(\{E(h)\}, \{F_0(h)\}, \{F_1(h)\}\) and \(\{F_2(h)\}\), which are functions of the radial distance \(r_j\) of the pseudo-atoms, and not of their triangular position:

\[
E(h) \approx 4\pi \int_0^{4\pi/3} \frac{1}{t} [\sin(t)/t]^2 \langle \nu_1(r) \rangle \text{ (II-6-5a)}
\]

\[
F_i(h) \approx (4\pi/3) \int_0^{4\pi/3} \frac{1}{t} \langle \cos(t - \sin(t)/t)^2 \rangle \langle \Delta \rho_i^2(r) \nu_1(r) \rangle \text{ (II-6-5b)}
\]

\[
t = \pi rh/D_m. \text{ (II-6-5c)}
\]

Thirdly, we take advantage of this simple \(r\) dependence to decompose the integrals of (II-6-5) into a sum of \(A\) spherical shells of radii \(r_{\lambda}\), over each of which the functions \(\langle \nu_1(r) \rangle\) and \(\langle \Delta \rho_i(r) \rangle\) are approximately constant:

\[
\langle \nu_1(r) \Delta \rho_i^2(r) \rangle = \sum_{\lambda=1}^{A} \gamma_{\lambda}(r) a_{\lambda} b_{\lambda}, \text{ (II-6-6a)}
\]

\[
\gamma_{\lambda}(r) = 1 \text{ if } r_{\lambda-1} \leq r \leq r_{\lambda} \text{ otherwise } \gamma_{\lambda}(r) = 0 \text{ (II-6-6b)}
\]

\[
\gamma_{\lambda}(r) = 1 \text{ if } r_{\lambda-1} \leq r \leq r_{\lambda} \text{ otherwise } \gamma_{\lambda}(r) = 0 \text{ (II-6-6b)}
\]
\[ a_z = 0; \forall z \]  \hspace{1cm} (II-6-6c)
\[ r_0 = 0; \quad r_\Lambda = D_m/2 \]  \hspace{1cm} (II-6-6d)
\[ a_z = 4\pi \int_0^{r_m^2} \gamma_\Lambda(r) \langle v_\Lambda(r) \rangle r^2 \, dr \]  \hspace{1cm} (II-6-6e)
\[ b_{\lambda, \nu} = \frac{4\pi}{a_z} \int_0^{r_m^2} \gamma_\Lambda(r) \langle \Delta \rho_\Lambda(r) \rangle r^2 \, dr. \]  \hspace{1cm} (II-6-6f)

Replacing (II-6-6) in (II-6-5) yields the following expressions for \( E(h) \) and \( F_\Lambda(h) \):

\[ E(h) \approx 4\pi \sum_{\lambda=1}^4 a_\lambda \int_0^{r_m^2} \gamma_\Lambda(r) [\sin t / t]^2 r^2 \, dr \]  \hspace{1cm} (II-6-7a)
\[ F_\Lambda(h) \approx \frac{4\pi}{3} \sum_{\lambda=1}^4 a_\lambda \gamma_\Lambda(r) \int [\cos t - \sin t / t]^2 r^2 \, dr, \]  \hspace{1cm} (II-6-7b)

namely

\[ E(h) = (D_m^2 / \pi^2 h^3) \sum_{\lambda=1}^4 (a_\lambda - a_{\lambda+1}) (\pi K_\lambda h - \sin \pi K_\lambda h) \]  \hspace{1cm} (II-6-8a)
\[ F_\Lambda(h) = (4D_m^2 / 3h) \sum_{\lambda=1}^4 (a_\lambda b_{\lambda, \nu} - a_{\lambda+1} b_{\lambda+1, \nu}) \times \left[ (\pi K_\lambda h + \sin \pi K_\lambda h) / 4 + (\cos \pi K_\lambda h - 1) / \pi K_\lambda h \right] \]  \hspace{1cm} (II-6-8b)
\[ K_\lambda = 2r_\Lambda / D_m. \]  \hspace{1cm} (II-6-8c)

Equations (II-6-6) are functions of the \((3\Lambda - 1)\) parameters \( \{r_\Lambda, a_\lambda, b_{\lambda, \nu}\} \).

The final problem is to determine the parameters \( \{r_\Lambda, a_\lambda, b_{\lambda, \nu}\} \) knowing \( X_{o\Lambda,h} \). Several constraints can be used for that purpose. The following five moments \( \{M_{\mu, \nu}\} \) [see (II-1-2)] can be expressed in terms of some of the parameters \( i_\Lambda(0) \), \( i_\Lambda'(0) \) and \( p_\nu(0) \):

\[ M_{0,0} = v_1 = [i_\Lambda(0)]^{1/2} \]  \hspace{1cm} (II-6-9a)
\[ M_{0,1} = 0 \]  \hspace{1cm} (II-6-9b)
\[ M_{0,1} = v_1 [3i_\Lambda'(0)/8\pi^2 i_\Lambda(0)] \]  \hspace{1cm} (II-6-9c)
\[ M_{1,0} = p_\nu(0) \]  \hspace{1cm} (II-6-9d)
\[ M_{0,2} = p_\nu(0). \]  \hspace{1cm} (II-6-9e)

These parameters \( i_\Lambda(0) \), \( i_\Lambda'(0) \) and \( p_\nu(0) \) are functions of \( X_\Lambda \) [Taupin & Luzzati, 1982]:

\[ i_\Lambda(0)/m = 2 \sum_{h=1}^\infty (-1)^{h+1} X_{o\Lambda,h} \]  \hspace{1cm} (II-6-10a)
\[ i_\Lambda'(0)/m = (8/3)D_m \sum_{h=1}^\infty (-1)^{h} [\pi^2 - 6h^2] X_{o\Lambda,h} \]  \hspace{1cm} (II-6-10b)
\[ p_\nu(0)/m = [4\pi/(3D_m^3)] \sum_{h=1}^\infty h^2 X_{o\Lambda,h} \]  \hspace{1cm} (II-6-10c)
\[ X_{o\Lambda,h} = X_{o\Lambda,h} \quad \text{if} \quad h \leq h_m \]  \hspace{1cm} (II-6-10d)

\[ X_{o\Lambda,h} = (h_m/h)^4 X_{o\Lambda,h} \quad \text{if} \quad h > h_m \]  \hspace{1cm} (II-6-10e)

([II-6-10e] corresponds to Porod’s law). Linking (II-6-9) and (II-6-10) together requires the knowledge of the number \( m \) of electrons in one particle. \( m \) can be determined knowing another parameter \( Q \), which is determined by thermodynamic experiments independently of the X-ray scattering study [Luzzati & Tardieu (1980); \( Q \) is defined in equation (28) of that paper]:

\[ m = Qi_\Lambda(0)/m. \]  \hspace{1cm} (II-6-11)

Additional constraints may be provided by the analysis of the autocorrelation functions (which are functions of \( X_\Lambda \)) and by the physical and chemical properties of the particle (see the example in §III).

A more radical alternative, discussed by Tardieu et al. (1976) for particle of quasi-spherical symmetry, is to determine approximate expressions of the functions \( \langle v_\Lambda(r) \rangle \) and \( \langle \Delta \rho_\Lambda(r) \rangle \) using \( i_\Lambda(h) \). If this is possible, then the functions \( E(h) \) and \( F_\Lambda(h) \) can be calculated directly using (II-6-5).

In any event, error analysis usually may be expected to be more sensitive to the nature of the problem than to its particular solution. This anticipation is confirmed empirically by the example of LDL discussed in §II: in this case the curves \( u \) vs \( w \) are indeed barely sensitive to the model used to describe the structure of the particle.

In conclusion, for practical purposes, (II-5-6) may take the form

\[ |\Sigma(u, w, X_\Lambda)| = |\Sigma|. \]  \hspace{1cm} (II-6-13)

which can be used to determine \( u \) and \( w \) when the experimental parameters \( \{X_{o\Lambda,h}\} \) and \( Z \) are known.

III. An illustration: a low-density serum lipoprotein (LDL)

The solution X-ray scattering study of a low-density serum lipoprotein extracted from Rhesus monkeys (Luzzati, Tardieu & Aggerbeck, 1979) provides a convenient illustration of the mathematical treatment of §II. This lipoprotein is a compact quasi-spherical particle, containing approximately 18% protein, 48% non-polar lipids (mainly cholesterol esters), 34% polar lipids; its molecular weight is \( 3.43 \times 10^6 \) daltons. The X-ray scattering experiments were performed at 294 K in water containing variable amounts of NaBr. The chemical composition, the data reduction process and the analysis of \( X_{o\Lambda,h} \) yield the values of the parameters
INFORMATION CONTENT IN SOLUTION SCATTERING STUDIES. II

Table 1. Some parameters relevant to the solution X-ray scattering study of LDL

\[ D_w = 305 \text{ Å}, \rho_1 = 0.3417 \text{ eÅ}^{-3} \text{ and } \{X_0, h\} \text{ are determined by the data reduction process for } h > h_{\alpha} = 27 \text{ the intensities are assumed to fulfill Porod's law, see (II-6-10)}; \]

\[ \rho_1(0)/m = -4.5 \times 10^{-3} \text{ eÅ}^{-3}, \rho_0(0)/m = 2.1 \times 10^{-2} \text{ eÅ}^{-3}, \rho_1(0)/m = 3.90 \times 10^{-1} \text{ eÅ}^{-3}. \]

Knowledge of these experimental parameters, and also the partial specific volume of LDL, yields the value of the following structural parameters (Luzzati et al., 1979; Luzzati & Tardieu, 1980):

\[ m = 1.83 \times 10^6 \text{ e}, R_e = 93.4 \text{ Å}, m_{2,1}/v_1 = 118 \text{ eÅ}^{-2}. \]

<p>|</p>
<table>
<thead>
<tr>
<th>$h$</th>
<th>$s_{i,j}(h)/m$</th>
<th>$s_{i,j}(h)/m$</th>
<th>$s_{i,j}(h)/m$</th>
<th>$s_{i,j}(h)/m$</th>
<th>$s_{i,j}(h)/m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0057</td>
<td>107.45</td>
<td>23.85</td>
<td>15</td>
<td>0.6024</td>
</tr>
<tr>
<td>2</td>
<td>1.9124</td>
<td>368.96</td>
<td>17.69</td>
<td>16</td>
<td>0.8294</td>
</tr>
<tr>
<td>3</td>
<td>3.7732</td>
<td>249.49</td>
<td>42.66</td>
<td>17</td>
<td>1.1073</td>
</tr>
<tr>
<td>4</td>
<td>2.3011</td>
<td>19.19</td>
<td>73.1</td>
<td>18</td>
<td>1.629</td>
</tr>
<tr>
<td>5</td>
<td>0.9255</td>
<td>7.04</td>
<td>85.8</td>
<td>19</td>
<td>0.8863</td>
</tr>
<tr>
<td>6</td>
<td>1.3000</td>
<td>20.80</td>
<td>48.8</td>
<td>20</td>
<td>0.6067</td>
</tr>
<tr>
<td>7</td>
<td>1.1353</td>
<td>40.34</td>
<td>230.5</td>
<td>21</td>
<td>0.4296</td>
</tr>
<tr>
<td>8</td>
<td>0.8346</td>
<td>1.68</td>
<td>259.2</td>
<td>22</td>
<td>0.3426</td>
</tr>
<tr>
<td>9</td>
<td>0.9255</td>
<td>0.12</td>
<td>150.3</td>
<td>23</td>
<td>0.2783</td>
</tr>
<tr>
<td>10</td>
<td>0.7860</td>
<td>1.12</td>
<td>104.2</td>
<td>24</td>
<td>0.2423</td>
</tr>
<tr>
<td>11</td>
<td>0.6341</td>
<td>2.33</td>
<td>113.1</td>
<td>25</td>
<td>0.2277</td>
</tr>
<tr>
<td>12</td>
<td>0.7632</td>
<td>0.30</td>
<td>71.13</td>
<td>26</td>
<td>0.2251</td>
</tr>
<tr>
<td>13</td>
<td>0.6816</td>
<td>-2.49</td>
<td>88</td>
<td>27</td>
<td>0.1901</td>
</tr>
<tr>
<td>14</td>
<td>0.5259</td>
<td>-2.01</td>
<td>77</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The two-shell model of LDL

The structure (see § II.6) is decomposed into two spherical shells of radii $r_1$ and $r_2$ and the functions $\langle v(r) \rho'(r) \rangle$ are assumed to take the form $a_2b_2$ over the first shell, $a_1b_1$ over the second, for $v = 0, 1, 2$ [see equations (II-6-6-8)]. The dimension of the model agree with the structure of LDL (see § III).

\[ \begin{bmatrix} a_1b_1 & a_2b_2 \end{bmatrix} \]

we compute the determinant $|\Sigma'|$ (A-1-8) and we solve (II-6-13). The final result is the curve $u$ vs $w$ plotted in Fig. 1.

Empirical tests show that the curves $u$ vs $w$ are

![Fig. 1. A few curves $u$ vs $w$ relevant to LDL (see § III). $u = \sigma_{pos}/n^{1/2}$, $w = \sigma_{char}/n^{1/2}$, $\sigma_{pos}$ and $\sigma_{char}$ being the variances of the position and chromatic parameters and $n$ the number of pseudo-atoms used in the representation [see (II-4-3c) and (II-4-3d)].](image)

<table>
<thead>
<tr>
<th>Curve</th>
<th>Degrees of freedom</th>
<th>Matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>$i(h)$ $i_0(h)$ $i_i(h)$ $i_{il}(h)$</td>
<td>Full</td>
</tr>
<tr>
<td>(a')</td>
<td>$i(h)$ $i_0(h)$ $i_i(h)$ $i_{il}(h)$</td>
<td>Diagonal terms only</td>
</tr>
<tr>
<td>(b)</td>
<td>$i(h)$ $i_0(h)$ $i_i(h)$</td>
<td>Full</td>
</tr>
<tr>
<td>(b')</td>
<td>$i(h)$ $i_0(h)$</td>
<td>Diagonal terms only</td>
</tr>
<tr>
<td>(c)</td>
<td>$i_0(h)$ $i_i(h)$ $i_{il}(h)$</td>
<td>Full</td>
</tr>
<tr>
<td>(c')</td>
<td>$i_0(h)$ $i_i(h)$</td>
<td>Diagonal terms only</td>
</tr>
<tr>
<td>(d)</td>
<td>$i_i(h)$ $i_{il}(h)$</td>
<td>Full</td>
</tr>
<tr>
<td>(d')</td>
<td>$i_i(h)$</td>
<td>Diagonal terms only</td>
</tr>
</tbody>
</table>

The arrows point to the asymptotic values of $u$ and $w$. Curve (a) corresponds to the interpretation of the variable contrast experiment within the framework of the invariant volume hypothesis. Curve (b) corresponds to the case in which the invariant volume hypothesis is forgone and only the curve at $\tilde{p} = p_1$ is invested with physical significance. Curves (c) and (d) do not correspond to any clear-cut physical problem (see text).
barely dependent upon the dimensions of the model, at least as long as these are physically plausible and fulfill the constraints of (II-6-9).

As expected, the variables $u$ and $w$ are inversely related: posting improved positional parameters $\{r_j\}$ has the effect of deteriorating the chromatic parameters $\{\rho_j\}$ and *vice versa*. Note also that as $w \to 0$ $u$ tends towards an asymptotic value, whereas $w$ increases indefinitely as $u \to 0$.

We also report in Fig. 1 the curves $u$ vs $w$ obtained when (II-6-13) is applied to submatrices relevant to only one of the three characteristic functions. In spite of the reservations mentioned in § II.5, Fig. 1 provides an interesting illustration of the relative sensitivity of each of the characteristic functions to the structural parameters. As expected, $i_v(h)$ is independent of $w$; besides, $i_p(h)$ is more sensitive than $i_{iv}(h)$ with regard to both the positional and the chromatic parameters. In keeping with this observation we may point out that the most accurate X-ray experiments on LDL are those performed at low NaBr concentration, near buoyancy ($\bar{p} \simeq \rho_1$); therefore, the determination of $i_v(h)$ involves a mild interpolation of the most accurate data, the determination of $i_p(h)$ is more dependent upon the less-accurate experiments at higher NaBr concentration, whereas $i_v(h)$ is defined by an extrapolation to $\bar{p} = \infty$.

We also plot in Fig. 1 the curves $u$ vs $w$ obtained using only the diagonal terms of the determinants $|\Sigma|$ and $|\Sigma^p|$ [see (II-5-8)]. Accuracy always deteriorates (namely $u$ and $w$ increase) with respect to the case in which the determinants are complete. This is an obvious consequence of the offset cross correlations being stronger in the real than in the virtual experiment (see §II.5). It may also be noted that the difference between the two cases is larger when the three characteristic functions are treated at the same time than when only one is analysed (Fig. 1). This latter observation is consistent with the presence of strong correlations between the three experimental characteristic functions, which probably are the consequence of the quasi-spherical symmetry of LDL [for strictly spherical symmetry $i_{ip}(h)$ is uniquely defined by $i_p(h)$ and $i_v(h)$, except for the sign, see Luzzati et al. (1976)].

### IV. Discussion

It is timely to emerge from the maze of the mathematical formalism and to express in plain words the substance of the problem involved. For the sake of clarity, we refer to the five steps of the mathematical treatment (§II.2 to II.6).

(a) **Data reduction:** parametrization in experimental space. The problem of specifying the degrees of freedom of the solution scattering experiment and of determining their most probable value and associated errors has been treated previously (Luzzati, 1980; Taupin & Luzzati, 1982). Suffice to stress here that these operations are heavily dependent upon the notions that the autocorrelation function has a finite support and that the asymptotic form of the intensity curves can be parametrized (we use Porod's law). Moreover, we have treated variable contrast experiments within the framework of the invariant volume hypothesis: it is worthwhile to stress that this hypothesis is optional (see §II.5 and the accompanying paper). Finally, the degrees of freedom of the scattering experiment are the values of the three characteristic functions at a finite number of points (Shannon's channels).

(b) **Representation of the structure:** parametrization in structural space. It is most common in X-ray scattering experiments to represent the electron density distribution in terms of spherically symmetric atomic components, whose shape is supposed to be known. In this case the degrees of freedom are the three coordinates of each atom and a few parameters specifying atomic number, shape of the electron density cloud and possibly molecular and/or lattice disorder. Clearly, this parametrization is suitable for high-resolution crystallographic analyses, not for low-resolution solution scattering studies. We chose, instead, to represent the structure in terms of sharp pseudo-atoms, whose number $n$ is smaller than the number of atoms; three degrees of freedom are required to specify the position of each pseudo-atom and one more to define its electron density. The choice of this particular representation is due to a large extent arbitrary: its main justification is expediency and mathematical simplicity; its main drawback - whose effects subside as $n$ increases - is the formal disagreement with Porod's law. Alternative representations could also be thought of [for example, expansion in spherical harmonics (Stuhrmann, 1970)]; in this case, though, the mathematical treatment should be revised and adapted to the new representation.

(c) **Introduction of a virtual experiment:** parametrization of the uncertainty of the structure parameters. We have adopted a stochastic model that consists in distorting the position and electron density of the pseudo-atoms away from their 'correct' values by a noise whose probability law is isotropic (with regard to position) and is the same for all the pseudo-atoms. The noise is characterized by the variances $\sigma_{pos}^2$ and $\sigma_{chr}^2$ ([II-4-3c] and [II-4-3d]). Subsequently, we have determined in reciprocal space the probability law of the intensities corresponding to that stochastic model: this law is a function of the variances and of the values of the structure parameters.

(d) **Comparison of the real and the virtual experiments:** The data reduction process yields, in reciprocal space, the probability law of the values of the three characteristic functions at Shannon's channels.
Similarly, the analysis of the virtual experiment leads to the determination of another probability law of the same variables. Our approach is based upon the interpretation of the uncertainty of the experimental observations as if it arose from the noise of the structure parameters; in other words, we strive to estimate the variances \( \sigma_{pos}^2 \) and \( \sigma_{chr}^2 \) by comparing, in reciprocal space, the two probability laws of the degrees of freedom, one corresponding to the real, the other to the virtual experiment. For that purpose we resort to the information, a parameter expressing in rigorous mathematical terms the probability of observing \textit{a posteriori} some variable with a particular probability law, when \textit{a priori} the probability law was expected to take a different form. The determination of the information requires the explicit expressions of both the \textit{a priori} and the \textit{a posteriori} probability laws: the latter is known, but not the former, which depends upon the statistical definition of the state of complete ignorance. The explicit formulation of the \textit{a priori} probability law involves some tricky mathematical problems, which we by-pass by assuming that the \textit{a priori} law is much flatter than the laws corresponding to both the real and the virtual experiments. Finally, the comparison of the two experiments amounts to assuming that the centres of the two probability laws coincide (in other words, that the 'correct' values of the structure parameters correspond to the most probable values of the observed intensities) and that the information takes the same value in the two cases. The result takes the form of an equation relating the normalized variances \( \sigma_{pos}^2/n, \sigma_{chr}^2/n \) of the structure parameters.

\( (e) \) Attenuation of the structure dependence. We mean in this work to assess the potential accuracy of a structure analysis before the problem is solved; more precisely, we seek an algorithm defining errors as a function of the value and associated uncertainty of the experimental intensities, independently of the (unknown) value of the structure parameters. We manage to eliminate the structure parameter dependence by introducing the assumption that the structure fulfills, at least to the first and coarse approximation, the requirements of spherical symmetry. We introduce the notation of (II-6-1) and (II-6-2) and remembering the definition

\[ T_j (s) = \frac{\sin 2 \pi r_j s}{2 \pi r_j s} = \exp \left( -2 \pi i (r_j - r_0) s \right) \]

we find that the expression for the sum \( \varphi \sum_{j=1}^{n} T_j (s) \) takes the form [see (II-3-2) and (II-3-1)]

\[ \varphi \sum_{j=1}^{n} T_j (s) = \left\langle \varphi \sum_{i=1}^{n} \exp(2 \pi i r_i s) \exp(2 \pi i r_j s) \right\rangle = \left\langle \exp(2 \pi i r_j s) \right\rangle \int v_i (r) \exp(-2 \pi i r_i s) \, dr \]

\[ = \left[ \sigma_i (s) + \delta_i (s) \right] \exp(2 \pi i r_j s) \]

Introducing, moreover, a hypothesis of quasi-spherical symmetry [namely (II-6-3)], we find that the partial sums of (4-1-2) take the form

\[ \varphi \sum_{i=1}^{n} T_j (s) \simeq \left[ \sigma_i (s) + \delta_i (s) \right] \exp(2 \pi i r_j s). \quad (A-1-3c) \]
This and a similar treatment of the other partial sums of \((A-1-2)\) lead to the following results:

\[
\varphi \sum_{i=1}^{n} T_j(s) \approx [\sigma_j(s) + \delta_j(s)] T_j(s) \quad (A-1-4a)
\]

\[
\varphi \sum_{i=1}^{n} \rho_i T_j(s) \approx [\sigma_j(s) + \delta_j(s)] T_j(s) \quad (A-1-4b)
\]

\[
\varphi \sum_{i=1}^{n} \varepsilon_j(s) \approx [\sigma_j(s) + \delta_j(s)] \varepsilon_j(s) \quad (A-1-4c)
\]

\[
\varphi \sum_{i=1}^{n} \rho_i \varepsilon_j(s) \approx [\sigma_j(s) + \delta_j(s)] \varepsilon_j(s) \quad (A-1-4d)
\]

\[
T_j(s) \equiv (\sin 2\pi r_j s) / 2\pi r_j s = \exp(2\pi i r_j s) \quad (A-1-4e)
\]

\[
\varepsilon_j(s) \equiv \Delta x_j \partial T_j(s) / \partial x_j + \Delta y_j \partial T_j(s) / \partial y_j
+ \Delta z_j \partial T_j(s) / \partial z_j. \quad (A-1-4f)
\]

Inserting equations \((A-1-4)\) into \((A-1-2)\), we get approximate expressions for \(\Delta X\):

\[
\Delta i_x(h) = 2\varphi \sum_{j=1}^{n} [A_{i_x}(h) T_j(h) \Delta \rho_j + \rho_j^* A_{i_x}(h) \varepsilon_j(h)] \quad (A-1-5a)
\]

\[
\Delta i_y(h) = -2\varphi \sum_{j=1}^{n} [A_{i_y}(h) T_j(h) \Delta \rho_j + \rho_j^* A_{i_y}(h) \varepsilon_j(h)]
\]

\[
+ A_{i_y}(h) \varepsilon_j(h)] \quad (A-1-5b)
\]

\[
\Delta i_z(h) = 2\varphi \sum_{j=1}^{n} A_{i_z}(h) \varepsilon_j(h) \quad (A-1-5c)
\]

\[
A_{i_x}(h) \equiv [\sigma_{i_x}(h) + \delta_{i_x}(h)] \quad (A-1-5d)
\]

\[
A_{i_y}(h) \equiv [\sigma_{i_y}(h) + \delta_{i_y}(h)]. \quad (A-1-5e)
\]

The determination of the elements of the matrix \(\Sigma^t\) can be pursued by replacing \((A-1-5)\) in \((A-1-1)\). The calculations involve the following steps and approximations:

(a) \(\Delta x_j, \Delta y_j, \Delta z_j, \Delta \rho_j\) are independent stochastic variables: the mathematical expectation of each of them and all their cross products is zero [see (II-4-3a) and (II-4-3b)];

(b) their variances \((e.g. \sigma([\Delta x_j]^2|P_z])\) are independent of \(j\) and related to the stochastic parameters \(u\) and \(w\) [see (II-4-3c) and (II-4-3d)];

(c) as a consequence of Shannon's theorem (see Taupin & Luzzati, 1982), and since \(D_m\) is the maximum chord of the particle, the overall contribution of products like \(T_j(h_1) T_j(h_2), T_j(h_1) \varepsilon_j(h_2), \varepsilon_j(h_1) \varepsilon_j(h_2)\) \((h_1 \neq h_2)\) are minimal, and the cross terms involving different lattice points \(h\) can be expected to be negligible (this point is easily verified by extending the calculations to these cross terms);

(d) as a consequence of the hypothesis of quasi-spherical symmetry (§II.6) we can introduce the following simplifications:

\[
A_{i_x}(h)^2 \approx \varepsilon_j(h) \quad (A-1-6a)
\]

\[
A_{i_y}(h) A_{i_z}(h) \approx -\frac{1}{2} \rho_j^* \varepsilon_j(h) \quad (A-1-6b)
\]

\[
A_{i_z}(h)^2 \approx \varepsilon_j(h); \quad (A-1-6c)
\]

(e) the parameter \(\varphi\) is a linear combination of the \(\{\varepsilon_{i,k}\}\) [see (II-6-9)]:

\[
\varphi = \frac{v_1}{n}. \quad (A-1-7)
\]

(f) the first non-zero terms of the expansion of the components of \(\Delta c\), namely \(\sigma[\Delta i(h)|P_z]\), are proportional to \(u^2\) and \(w^2\); therefore, their products can be neglected (see \((A-1-2)\)).

The final expressions of the non-zero elements of the matrix \(\Sigma^t\) are

\[
\Sigma^t_{i_x(i_x),i_y(i_y)} = 4(v_1/m)[\sigma_{i_x} E(h) u^2 + \sigma_{i_y} F_0(h) u^2] \quad (A-1-8a)
\]

\[
\Sigma^t_{i_x(i_x),i_z(i_z)} = 4(v_1/m)[\sigma_{i_y} E(h) u^2 + \sigma_{i_z} F_0(h) u^2]
\]

\[
+ \sigma_{i_z} F_0(h) u^2 \quad (A-1-8b)
\]

\[
\Sigma^t_{i_y(i_y),i_z(i_z)} = 2(v_1/m) [\sigma_{i_x} E(h) w^2 + \sigma_{i_z} F_0(h) w^2]
\]

\[
+ \sigma_{i_z} F_0(h) w^2 \quad (A-1-8c)
\]

\[
\Sigma^t_{i_x(i_x),i_y(i_y),i_z(i_z)} = 2(v_1/m) [-2 \sigma_{i_x} F_1(h) + \sigma_{i_y} F_0(h)] u^2 \quad (A-1-8d)
\]

\[
\Sigma^t_{i_y(i_y),i_z(i_z)} = 2(v_1/m) [-2 \sigma_{i_x} F_1(h) + \sigma_{i_y} F_0(h)] u^2 \quad (A-1-8e)
\]

\[
\Sigma^t_{i_x(i_x),i_y(i_y),i_z(i_z)} = 2(v_1/m) [-2 \sigma_{i_x} F_1(h) + \sigma_{i_y} F_0(h)] u^2 \quad (A-1-8f)
\]

where

\[
E(h) = \varphi \sum_{j=1}^{n} T_j^2(h) \quad (A-1-9a)
\]

\[
u^2 F_0(h) = \varphi \sum_{j=1}^{n} \rho_j^* \varepsilon_j^2(h). \quad (A-1-9b)
\]

Remembering the definitions for \(T_j(h)\) \((A-1-4e)\), \(\varepsilon_j(h)\) \((A-1-4f)\) and \(r_j\):

\[
r_j = (x_j^2 + y_j^2 + z_j^2)^{1/2}. \quad (A-1-10)
\]

we find that the expressions for \(E(h)\) and \(F(h)\) take the form of \((II-6-5)\). Note that these functions depend upon the radial positions \(r_j\) and the weights \(\rho_j\), but not upon the angular components of \(r_j\). We discuss in §II.6 the problem of expressing these functions in terms of the experimental parameters \(X_0\).

**References**


