

Probabilistic Inference for Scattering Data Analysis

Chi-Huan Tung

June 24, 2025

This document presents a unified framework for using probabilistic inference to extract physically meaningful quantities from experimental measurements. The approach is grounded in linear algebra, where matrix operations are employed to systematically propagate uncertainties and resolve latent variables from noisy or incomplete data. Particular emphasis is placed on cases where the target quantity is not directly accessible through measurements, but can be inferred through well-defined forward or inverse linear operators. To illustrate the versatility and broad applicability of this strategy, we present several case studies based on real-world challenges encountered in scattering experiments, including Small-Angle Neutron Scattering (SANS), Neutron Spin Echo (NSE), and X-ray Photon Correlation Spectroscopy (XPCS).

1 Regression Analysis

Regression analysis provides a systematic framework for estimating unknown physical parameters by aligning experimental measurements with predictions generated by a parameterized model. This process is typically formulated as an optimization problem:

$$\Lambda = \underset{\Lambda}{\operatorname{argmin}} \mathcal{L}(X, \Lambda), \quad (1)$$

where Λ denotes the model parameters, and the loss function $\mathcal{L}(X, \Lambda)$ quantifies the discrepancy between the measured data $Y(X)$ and the model output

$f(X, \Lambda)$:

$$\mathcal{L}(X, \Lambda) = \sum_X \|Y(X) - f(X, \Lambda)\|^2. \quad (2)$$

The parameters Λ are iteratively adjusted through a curve-fitting procedure to minimize the loss function. This enables two key outcomes: (1) the inferred parameters Λ may correspond to physically meaningful quantities, and (2) the model $f(X, \Lambda)$ can interpolate or smooth the data, allowing estimation in regions where direct measurements are unavailable and enhancing numerical robustness.

In practice, however, the functional form f that relates a low-dimensional parameter set Λ to the observable Y is often unknown, inaccessible, or too complex to specify analytically. Moreover, experimental observations are inevitably corrupted by noise. As a result, the idealized relationship

$$Y = f(X, \Lambda) \quad (3)$$

is generally violated. A more realistic formulation accounts for measurement uncertainty via an additive noise term:

$$Y = f(X, \Lambda) + \eta(X), \quad (4)$$

where $\eta(X)$ denotes the noise, typically modeled as a random variable informed by the statistical properties of the measurement process.

Given these challenges, the goal of data analysis becomes the recovery of a physically meaningful estimate of the latent parameters Λ from noisy observations Y , under the following conditions: (1) the relationship between Λ and Y can be approximated by a linear operator; and (2) the observation noise is Gaussian-distributed and characterized by a known covariance structure.

A principled and flexible solution is to adopt a **probabilistic inference framework**, which enables uncertainty-aware estimation by incorporating prior knowledge about the system and explicitly modeling the effects of measurement noise.

2 Probabilistic Description of Regression Loss

To motivate the probabilistic framework, we begin with a simple example: the one-dimensional **Gaussian (normal) distribution**. A scalar random

variable $y \in \mathbb{R}$ is said to follow a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, denoted as

$$y \sim \mathcal{N}(\mu, \sigma^2), \quad (5)$$

if its probability density function is given by:

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right). \quad (6)$$

This distribution describes the likelihood of observing different values of y , assuming that values closer to the mean μ are more probable, and the spread is governed by the variance σ^2 .

This concept generalizes naturally to higher dimensions. The **multivariate normal distribution** is a cornerstone of probabilistic modeling, particularly in the context of regression with Gaussian noise. A random vector $\mathbf{y} \in \mathbb{R}^N$ is said to follow a multivariable normal distribution with mean $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance matrix $\Sigma \in \mathbb{R}^{N \times N}$, written as

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad (7)$$

if its probability density function (PDF) is given by:

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{N/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})\right). \quad (8)$$

The mean vector $\boldsymbol{\mu}$ represents the expected value of the random vector \mathbf{y} , while the covariance matrix Σ captures both the variances of individual components and the correlations between them. The quantity

$$\|\mathbf{y} - \boldsymbol{\mu}\|_{\Sigma^{-1}}^2 = (\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}) \quad (9)$$

is known as the **Mahalanobis distance**, which generalizes the notion of distance by incorporating the geometry induced by the covariance matrix. $\|\cdot\|_{\Sigma^{-1}}$ denotes the Mahalanobis norm:

$$\|\mathbf{v}\|_{\Sigma^{-1}}^2 = \mathbf{v}^\top \Sigma^{-1} \mathbf{v}. \quad (10)$$

Level sets of constant Mahalanobis distance form **ellipsoidal contours** in multivariate space, where the orientation and shape of each ellipsoid reflect

the underlying covariance structure: longer axes indicate directions of greater variance, while shorter axes correspond to directions of tighter constraint.

This formulation naturally extends to modeling noisy measurements in regression, where the observations Y are treated as Gaussian-distributed random variables around a deterministic model prediction. The multivariable normal distribution provides a compact and mathematically convenient way to encode both measurement uncertainty and statistical dependence across different observations. Assuming the noise term $\eta(X)$ follows a Gaussian distribution with zero mean and known (or estimable) variance, Eqn. (4) becomes:

$$Y = f(X, \Lambda) + \eta(X), \quad \eta(X) \sim \mathcal{N}(0, \Sigma), \quad (11)$$

where Σ is the noise covariance matrix. Under this assumption, the **likelihood** of observing data Y given parameters Λ is:

$$p(Y | X, \Lambda) = \frac{1}{Z} \exp \left(-\frac{1}{2} \|Y - f(X, \Lambda)\|_{\Sigma^{-1}}^2 \right), \quad (12)$$

where Z is the normalization constant.

This probabilistic perspective reveals that minimizing the squared loss function:

$$\mathcal{L}(X, \Lambda) = \|Y - f(X, \Lambda)\|_{\Sigma^{-1}}^2, \quad (13)$$

is equivalent to performing **maximum likelihood estimation (MLE)** of Λ under Gaussian noise.

The loss function in Eqn (13) simply quantifies the disagreement between the measured data Y and the model prediction $f(X, \Lambda)$. However, if prior knowledge or assumptions about the parameters Λ are available before observing the data Y , it is appropriate to incorporate this information into the inference process. In such cases, we introduce a **prior** distribution over Λ :

$$p(\Lambda) = \frac{1}{Z'} \exp \left(-\frac{1}{2} \|\Lambda - \Lambda_0\|_{K^{-1}}^2 \right), \quad (14)$$

Given the likelihood function $p(Y | X, \Lambda)$ and the prior distribution $p(\Lambda)$, **Bayes' theorem** provides a principled way to update our beliefs about the parameters Λ after observing the data Y . The resulting distribution is known as the **posterior**:

$$p(\Lambda | X, Y) = \frac{p(Y | X, \Lambda) p(\Lambda)}{p(Y | X)}, \quad (15)$$

where the denominator $p(Y | X)$, known as the marginal likelihood or model evidence, acts as a normalization constant to ensure the posterior integrates to one. It is independent of Λ and does not affect parameter estimation.

A common strategy is to compute the **maximum a posteriori** (MAP) estimate, which identifies the most probable value of Λ under the posterior distribution:

$$\Lambda_{\text{MAP}} = \underset{\Lambda}{\operatorname{argmax}} p(\Lambda | X, Y). \quad (16)$$

Taking the negative logarithm of the posterior and ignoring constants independent of Λ , the MAP estimation reduces to the minimization of a regularized loss function:

$$\mathcal{L}(X, \Lambda) = \|Y - f(X, \Lambda)\|_{\Sigma^{-1}}^2 + \|\Lambda - \Lambda_0\|_{K^{-1}}^2. \quad (17)$$

3 The Role of Data Smoothness in Enhancing Information Throughput

In this section, we explain the benefits of data smoothness from the perspective of Shannon’s noisy channel theorem [1]. In scattering experiments, particularly those with limited counting statistics (e.g., neutron or photon-limited measurements), data continuity provides an underappreciated but powerful source of redundancy that can be harnessed to improve the effective quality of the measurement. This smoothness not only reduces the apparent noise but also enables the construction of priors distributions as discussed in Eqn. (14), which serve as regularizers in probabilistic inference and enhance the stability and reliability of parameter estimation.

Shannon’s noisy channel theorem describes the fundamental trade-off between three quantities:

- Message size M : the amount of information being transmitted or extracted,
- Repetition rate n : the number of statistically independent samples or trials,
- Error probability p_{err} : the chance of misinterpreting or misestimating the message.

Two dual interpretations of this relationship are especially relevant for experimental data:

1. For fixed message size M , increasing the number of repetitions n reduces the error probability p_{err} .
2. For fixed error tolerance p_{err} , increasing n allows for a larger message size M , i.e., more information can be accessed from the same measurement.

In the context of experimental data, smoothness in the measured function, such as the continuity of $I(Q)$ in small-angle scattering, acts as a form of implicit repetition. If the function varies slowly over a characteristic length scale λ , then nearby measurements $I(Q_i)$ and $I(Q_{i+1})$, taken at spacing $\delta \ll \lambda$, are statistically correlated and effectively constitute repeated observations of the same underlying value.

Formally, for any $\epsilon > 0$, if $|Q_2 - Q_1| < \lambda$, then $|I(Q_2) - I(Q_1)| < \epsilon$. This implies that within a local window of size λ , we obtain multiple statistically similar observations of the same signal. The effective repetition rate is increased by a factor:

$$\mu \sim \frac{\lambda}{\delta}, \quad (18)$$

where δ is the sampling interval. This results in a reduction of effective noise and an enhancement in the information that can be extracted from the measurement. Using this interpretation, the effect of smoothness on measurement precision can be quantified as:

- Fixing message size M , the error probability decreases exponentially with the effective repetition:

$$p_{\text{err}} \sim \exp(-k\mu), \quad k > 0, \quad (19)$$

- Fixing p_{err} , the log-accessible message size increases with μ :

$$\log M > 0 \sim \mu. \quad (20)$$

These relations encapsulate a central insight: smoothness serves as an information amplifier. The statistical performance of a measurement is not determined solely by physical parameters such as flux or detector resolution. Instead, it is fundamentally shaped by the structure of the signal itself.

Smoothness introduces an effective repetition in the observed data, enabling inference algorithms to “see more than what is directly measured.”

Inference algorithms that incorporate smooth priors naturally leverage this structure. By encoding continuity into the prior distribution, these methods align the inference process with the intrinsic regularity of the physical system. This alignment enhances robustness to noise and improves reconstruction accuracy, particularly in low-flux or sparsely sampled experiments.

4 Applications in Scattering Data Analysis

To illustrate the practical utility and generality of the proposed probabilistic inference framework, we present three representative case studies drawn from experimental scattering techniques. Each case study is characterized by distinct physical observables, inference models, and experimental data formats. Despite these differences, all cases share a unified structure grounded in Bayesian inference, used to estimate latent physical quantities from noisy or incomplete measurements.

For each scenario, we outline the formulation of the inference problem by specifying:

- the model function $f(X, \Lambda)$ that links latent parameters Λ to observed data,
- the definition and physical interpretation of the parameter space Λ ,
- the form of the prior distribution $p(\Lambda)$, encoding assumptions such as smoothness or sparsity,
- the structure of the observation, noise and its covariance Σ .

We then derive the corresponding maximum a posteriori (MAP) estimator for the inferred quantity, and highlight the computational and physical implications of each formulation.

4.1 Case I: Small Angle Scattering (SAS)

The physical quantity measured in small-angle scattering (SAS) experiments is the scattering intensity $I(Q)$, where Q denotes the magnitude of the momentum transfer. Since $I(Q)$ corresponds to the Fourier transform of the

autocorrelation function of the spatial distribution of scattering length density (e.g., mass or electron density), SAS provides a powerful means to probe the morphology of materials at nanometer to micrometer length scales. This makes it particularly valuable for investigating complex soft matter systems, polymers, colloids, and biological assemblies.

Despite its utility, SAS techniques suffer from inherent limitations in photon or neutron flux. For neutron scattering, source intensity improvements are extremely challenging: increasing flux typically requires major upgrades to accelerator or reactor-based facilities, which are costly and span decades of development. In the case of X-ray scattering, although laboratory and synchrotron sources can provide relatively high fluxes, the high energy of X-ray photons at equivalent wavelengths imposes practical constraints due to radiation damage and exposure limits. As a result, total exposure time is often restricted, limiting the available data quality and quantity.

These limitations motivate the use of probabilistic inference to enhance the extraction of physical information from sparse or noisy $I(Q)$ measurements. In this case study, we demonstrate how prior knowledge of material structure and smoothness constraints can be incorporated into a Bayesian framework to infer reliable scattering profiles and underlying structural parameters, even when data are limited.

The objective here is to reconstruct the true scattering intensity $I(Q)$ from experimentally measured data $I^{\text{Exp}}(Q)$, which is corrupted by random noise. We assume the measurement model follows an additive Gaussian noise model:

$$I^{\text{Exp}}(Q) = I(Q) + \eta(Q), \quad (21)$$

where $\eta(Q) \sim \mathcal{N}(0, \Sigma)$ represents a zero-mean Gaussian noise process with known or estimated covariance structure Σ . This defines the likelihood function:

$$p(I^{\text{Exp}} | I) = \frac{1}{Z} \exp \left(-\frac{1}{2} \|I^{\text{Exp}} - I\|_{\Sigma^{-1}}^2 \right), \quad (22)$$

where I is the latent noise-free intensity vector evaluated on the same set of Q -values as the data.

Since the transformation from parameter to observable is direct, we adopt the identity map:

$$f(Q, \Lambda) = \Lambda(Q), \quad (23)$$

and treat the unknown function $\Lambda(Q) = I(Q)$ as the inference target.

To encode smoothness in $I(Q)$, we place a multivariable Gaussian prior over Λ , assuming that the covariance between two intensity values at Q and Q' is given by a kernel function:

$$\text{Cov}(I(Q), I(Q')) = K(Q, Q'), \quad (24)$$

where $K(Q, Q')$ is a symmetric, positive-definite function encoding expected spatial correlations. A common choice is the squared exponential (Gaussian) kernel:

$$K(Q, Q') = \sigma^2 \exp\left(-\frac{(Q - Q')^2}{2\ell^2}\right), \quad (25)$$

where σ^2 controls the signal variance and ℓ defines the characteristic correlation length scale. This suppresses sharp fluctuations because such variations would require nearby points to differ significantly, which would be unlikely under a prior that assumes strong correlations (high covariance) between them, allowing robust reconstruction of $I(Q)$ even in the presence of noise or limited data.

In the following, we adopt matrix notation for clarity and generality. Let $\mathbf{I} \in \mathbb{R}^N$ denote the vector of noise-free scattering intensities evaluated at discrete momentum transfer values Q_1, \dots, Q_N , and let $\mathbf{I}^{\text{Exp}} \in \mathbb{R}^N$ denote the corresponding experimental measurements. We also define $\mathbf{K} \in \mathbb{R}^{N \times N}$ as the prior covariance matrix and $\mathbf{\Sigma} \in \mathbb{R}^{N \times N}$ as the noise covariance matrix.

- **Prior:** The smoothness of the latent scattering intensity \mathbf{I} is encoded by a multivariable normal distribution with zero mean and covariance \mathbf{K} :

$$p(\mathbf{I}) = \frac{1}{Z'} \exp\left(-\frac{1}{2} \mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I}\right). \quad (26)$$

- **Likelihood:** The measured data \mathbf{I}^{Exp} are modeled as Gaussian-distributed around the true intensity with noise covariance $\mathbf{\Sigma}$:

$$p(\mathbf{I}^{\text{Exp}} | \mathbf{I}) = \frac{1}{Z} \exp\left(-\frac{1}{2} (\mathbf{I}^{\text{Exp}} - \mathbf{I})^\top \mathbf{\Sigma}^{-1} (\mathbf{I}^{\text{Exp}} - \mathbf{I})\right). \quad (27)$$

- **Objective:** Maximizing the posterior distribution $p(\mathbf{I} | \mathbf{I}^{\text{Exp}})$ is equivalent to minimizing the negative log-posterior. This leads to the following objective function:

$$\mathcal{L}(\mathbf{I}) = (\mathbf{I}^{\text{Exp}} - \mathbf{I})^\top \mathbf{\Sigma}^{-1} (\mathbf{I}^{\text{Exp}} - \mathbf{I}) + \mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I}. \quad (28)$$

The first term in the objective function enforces agreement with the observed data (data fidelity), while the second term penalizes non-smooth behavior in \mathbf{I} according to the prior. Minimizing $\mathcal{L}(\mathbf{I})$ thus yields the maximum a posteriori (MAP) estimate of the scattering profile.

Owing to the quadratic form of $\mathcal{L}(\mathbf{I})$, the MAP estimate can be obtained analytically by solving a linear system. This approach is called *Gaussian Process Regression* (GPR) [2, 3, 4, 5, 6] because it stems from the assumption that the unknown function \mathbf{I} is drawn from a multivariate Gaussian distribution (a Gaussian process prior), while the observation model is linear with additive Gaussian noise. Setting the gradient of the objective with respect to \mathbf{I} to zero leads to the following equation:

$$(\boldsymbol{\Sigma}^{-1} + \mathbf{K}^{-1}) \mathbf{I}^{\text{MAP}} = \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}. \quad (29)$$

Solving this system yields the closed-form solution for the maximum a posteriori (MAP) estimate:

$$\mathbf{I}^{\text{MAP}} = (\boldsymbol{\Sigma}^{-1} + \mathbf{K}^{-1})^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}, \quad (30)$$

where \mathbf{K} is the prior covariance and $\boldsymbol{\Sigma}$ is the noise covariance matrix.

Alternatively, applying the Woodbury matrix identity¹, the solution can be expressed in a numerically advantageous form:

$$\mathbf{I}^{\text{MAP}} = \mathbf{K} (\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{I}^{\text{Exp}}. \quad (31)$$

Throughout this manuscript, we refer to Eqn. (30) as the **Forward Form**, as it arises directly from minimizing the objective function via forward variational calculus. In contrast, Eqn. (31) is termed the **Dual Form**, as it shifts

¹The dual form $\mathbf{I}^{\text{MAP}} = \mathbf{K}(\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{I}^{\text{Exp}}$ is equivalent to the forward form $(\boldsymbol{\Sigma}^{-1} + \mathbf{K}^{-1})^{-1} \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}$ by the Woodbury identity:

$$(\boldsymbol{\Sigma}^{-1} + \mathbf{K}^{-1})^{-1} = \mathbf{K} - \mathbf{K}(\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{K}.$$

Substituting this yields

$$\begin{aligned} \mathbf{I}^{\text{MAP}} &= (\mathbf{K} - \mathbf{K}(\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{K}) \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}} \\ &= \mathbf{K} [\mathbf{I}_N - (\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{K}] \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}} \\ &= \mathbf{K}(\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{I}^{\text{Exp}}, \end{aligned}$$

since the relation $\mathbf{I}_N - (\boldsymbol{\Sigma} + \mathbf{K})^{-1} \mathbf{K} = (\boldsymbol{\Sigma} + \mathbf{K})^{-1} \boldsymbol{\Sigma}$ can also be obtained by the Woodbury formula. \mathbf{I}_N denotes the identity matrix.

computation to the data space through the use of the Woodbury identity. The relative merits of these formulations in terms of computational efficiency, scalability, and interpretability will be discussed in detail in later sections. Both expressions highlight how the MAP estimate reflects a trade-off between two sources of information: fidelity to the measured data, governed by the noise precision Σ , and smoothness imposed by the prior, governed by \mathbf{K} .

This interpretation that prior structure can be leveraged to enhance the inference of latent quantities is not limited to static measurements. It extends naturally to time-resolved scattering experiments, where the underlying dynamics exhibit smooth and often decaying behavior.

4.2 Case II: Neutron Spin Echo (NSE)

The NSE experiment measures both spatial and temporal correlations, represented as $S(Q, t)$. In most liquids and soft matter systems, the intermediate scattering function decays monotonically in time due to relaxation processes. To effectively capture this behavior, we adopt a Prony series expansion [7, 8, 9, 10]:

$$S(Q, t) = \sum_{n=1}^N A(Q, \tau_n) e^{-t/\tau_n}, \quad (32)$$

where $A(Q, \tau_n)$ denotes the expansion coefficients and τ_n are fixed characteristic relaxation times. In matrix notation, this formulation can be expressed as a linear operation $\mathbf{C}\mathbf{A}$, where $\mathbf{C} \in \mathbb{R}^{M \times N}$ is the Prony basis matrix evaluated at measurement times t_a , and $\mathbf{A} \in \mathbb{R}^{L \times N}$ contains the expansion coefficients at inference positions Q_i .

However, in practical settings, the measurement coordinates (Q_a, t_a) do not coincide with the coordinates at which inference is desired, (Q_i, t_m) . To reconcile this mismatch, we introduce a row-normalized spatial weight matrix $\mathbf{W} \in \mathbb{R}^{M \times L}$, which linearly projects the latent coefficients defined at Q_i to the observation coordinates Q_a . The weights are defined using a Gaussian kernel:

$$W_{ai} = \frac{\exp\left(-\frac{|Q_a - Q_i|^2}{2\mu^2}\right)}{\sum_{j=1}^L \exp\left(-\frac{|Q_a - Q_j|^2}{2\mu^2}\right)}, \quad (33)$$

where μ is a spatial correlation length scale that controls the influence radius of each inference point. The combined spatial-temporal transformation from

the coefficient tensor A_{in} to the observations S_a^{Exp} is then expressed through the third-order transformation tensor:

$$G_{a,in} = C_{an} W_{ai}, \quad (34)$$

$$S_a^{\text{Exp}} = C_{an} W_{ai} A_{in} + \eta_a, \quad (35)$$

which maps the (Q_i, τ_n) -indexed latent space to observed space (Q_a, t_a) .² This compact transformation forms the core of the inference model and enables a flexible yet physically grounded projection of structural dynamics onto the experimental coordinate system.

In this formulation, the model function $f(X, \Lambda)$ represents the predicted scattering signal at experimental coordinates $X = (Q_a, t_a)$, given a set of latent expansion coefficients $\Lambda = A_{in}$, which encode the dynamical properties of the system over a predefined basis of relaxation times τ_n . The input variable X consists of the observed spatiotemporal measurement coordinates, while the output $f(X, \Lambda)$ is obtained by applying the transformation tensor $G_{a,in}$ to A_{in} . This tensor contraction effectively propagates the latent Prony coefficients, defined at inference coordinates (Q_i, τ_n) , to the detector coordinates (Q_a, t_a) . Thus, f encapsulates the forward model linking latent dynamics to measurable quantities, and Λ denotes the parameter space over which inference is performed.

- **Prior:** The latent Prony coefficients A_{in} are assigned a multivariate Gaussian prior with zero mean and structured covariance $K_{in,jq}$, which enforces smoothness across spatial coordinates and independence across exponential components:

$$p(A) = \frac{1}{Z'} \exp \left(-\frac{1}{2} A_{in} K_{in,jq}^{-1} A_{jq} \right), \quad \text{with } K_{in,jq} = k_{ij} \delta_{nq}. \quad (36)$$

The spatial kernel is defined as:

$$k_{ij} = \exp \left(-\frac{|Q_i - Q_j|^2}{2\lambda^2} \right), \quad (37)$$

²We adopt the `numpy.einsum` style notation [11] throughout this section. This convention provides a flexible and readable representation of tensor contractions, which is particularly useful when describing operations over mixed spatial and temporal coordinates. Unlike traditional Einstein summation notation [12], in which repeated indices always imply summation, the `einsum` style allows repeated indices to represent either summation or element-wise alignment, depending on the context and output dimensionality. This distinction enables compact representations of complex expressions while preserving the underlying structure of the transformation.

where λ is a length-scale hyperparameter controlling the decay of spatial correlations.

- **Likelihood:** The measured NSE data S_a^{Exp} are modeled as noisy observations of the forward-transformed coefficients through $G_{a,in}$, with additive Gaussian noise $\eta_a \sim \mathcal{N}(0, \Sigma_{ab})$:

$$p(S^{\text{Exp}} | A) = \frac{1}{Z} \exp \left(-\frac{1}{2} (S_a^{\text{Exp}} - G_{a,in} A_{in}) \Sigma_{ab}^{-1} (S_b^{\text{Exp}} - G_{b,jq} A_{jq}) \right), \quad (38)$$

where Σ_{ab} is the noise covariance matrix, often taken to be diagonal based on pointwise uncertainty estimates from experimental data.

- **Objective:** Maximizing the posterior distribution $p(A | S^{\text{Exp}})$ is equivalent to minimizing the negative log-posterior. This leads to the following regularized least-squares objective function:

$$\mathcal{L}(A) = (S_a^{\text{Exp}} - G_{a,in} A_{in}) \Sigma_{ab}^{-1} (S_b^{\text{Exp}} - G_{b,jq} A_{jq}) + A_{in} K_{in,jq}^{-1} A_{jq}. \quad (39)$$

Following the procedure described in the previous section, the MAP estimation of the Prony coefficients can be expressed as either a forward-form or a dual-form solution, depending on which matrix inversion is more computationally convenient.

- **Forward Form:** The forward form solves the normal equation in coefficient space:

$$A_{in}^{\text{MAP}} = (G_{a,in} \Sigma_{ab}^{-1} G_{b,jq} + K_{in,jq}^{-1})_{in,jq}^{-1} G_{b,jq} \Sigma_{ab}^{-1} S_a^{\text{Exp}}. \quad (40)$$

This formulation requires inversion of a matrix in coefficient space of size $(L \cdot N) \times (L \cdot N)$.

- **Dual Form:** Alternatively, using the Woodbury identity, the MAP estimate can be expressed in dual form, where the matrix size to be inverted is $M \times M$. This avoids direct inversion in the high-dimensional coefficient space:

$$A_{in}^{\text{MAP}} = K_{in,jq} G_{b,jq} (G_{a,in} K_{in,jq} G_{b,jq} + \Sigma_{ab})_{ab}^{-1} S_a^{\text{Exp}}, \quad (41)$$

Other scattering techniques can yield data structures and inference models that resemble those discussed in the NSE case. An example is **contrast variation small-angle neutron scattering (CV-SANS)**, where multiple scattering profiles $I(Q)$ are measured from chemically identical samples prepared with different solvent contrasts. In such systems, the total scattering function is expressed as a weighted linear combination of partial structure factors $S_{ij}(Q)$, determined by the scattering length density (SLD) contrasts $\Delta\rho_i$ between solutes and solvent.

For a three-component system in contrast variation SANS, the measurement equation at each scattering vector Q_i can be expressed as a linear regression model over the contrast index γ_m [13]:

$$\mathbf{I}_{im}^{\text{Exp}} = \mathbf{M}_{mn}\mathbf{S}_{ni} + \boldsymbol{\eta}_{im}, \quad (42)$$

where $\mathbf{I}_{im}^{\text{Exp}} \in \mathbb{R}^{L \times N}$ denotes the measured scattering intensities at L distinct Q_i values across $N \geq 3$ contrast conditions γ_m . The contrast matrix $\mathbf{M} \in \mathbb{R}^{N \times 3}$ is defined by the known SLD differences:

$$\mathbf{M}_{mn} = \begin{bmatrix} (\Delta_1\rho_1)^2 & (\Delta_1\rho_2)^2 & 2\Delta_1\rho_1\Delta_1\rho_2 \\ \vdots & \vdots & \vdots \\ (\Delta_N\rho_1)^2 & (\Delta_N\rho_2)^2 & 2\Delta_N\rho_1\Delta_N\rho_2 \end{bmatrix}, \quad (43)$$

and $\mathbf{S}_{ni} \in \mathbb{R}^{3 \times L}$ is the matrix of unknown partial structure factors at each Q_i , with rows corresponding to $S_{11}(Q_i)$, $S_{22}(Q_i)$, $S_{12}(Q_i)$. The additive noise term $\boldsymbol{\eta}_{im}$ accounts for measurement uncertainty across combinations of contrast condition and scattering vector.

This formulation mirrors the tensor contraction structure introduced in Eqn. (35) for the NSE case. Now, assume each row $\mathbf{S}_n = [\mathbf{S}_{n1}, \dots, \mathbf{S}_{nL}] \in \mathbb{R}^L$ follows a zero-mean multivariate normal distribution with a smooth kernel prior:

$$p(\mathbf{S}_{ni}) \propto \exp\left(-\frac{1}{2}\mathbf{S}_{ni}\mathbf{K}_{ij}^{-1}\mathbf{S}_{nj}\right), \quad (44)$$

where the prior covariance $\mathbf{K} \in \mathbb{R}^{L \times L}$ is given by:

$$\mathbf{K}_{ij} = \exp\left(-\frac{(Q_i - Q_j)^2}{2\lambda^2}\right). \quad (45)$$

We assume additive Gaussian noise with independent variance for each measurement:

$$\boldsymbol{\eta}_{im} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{im}). \quad (46)$$

The MAP estimate of the partial structure factors $\mathbf{S} \in \mathbb{R}^{3 \times L}$ can then be obtained by solving the regularized least-squares problem:

$$\mathbf{S}^{\text{MAP}} = (\mathbf{M}^\top \Sigma^{-1} \mathbf{M} + \mathbf{K}^{-1})^{-1} \mathbf{M}^\top \Sigma^{-1} \mathbf{I}^{\text{Exp}}, \quad (47)$$

or, equivalently, using the Woodbury matrix identity:

$$\mathbf{S}^{\text{MAP}} = \mathbf{K} \mathbf{M}^\top (\mathbf{M} \mathbf{K} \mathbf{M}^\top + \Sigma)^{-1} \mathbf{I}^{\text{Exp}}. \quad (48)$$

Such a formulation for indirect inference can be generalized to any scenario where the connection between observations and latent physical quantities is represented by a linear operator. This structure encompasses a broad class of methods that employ basis expansions to regularize noisy measurements. For instance, consider the case of **two-dimensional scattering spectra expressed in polar coordinates**, $I(Q, \phi)$. At fixed Q , the angular dependence $I_Q(\phi)$ can be represented by a truncated Fourier series:

$$I_Q(\phi) = \sum_n A_n(Q) \phi_n(\phi),$$

where $\phi_n(\phi)$ denotes the angular basis functions. This decomposition yields a linear transformation structure similar to Eqns. (35) and (42), where the coefficients $A_n(Q)$ are inferred from experimental data. In addition to the angular basis, smoothness along the radial Q -direction can be incorporated through a Gaussian process prior with covariance matrix \mathbf{K} , which promotes continuity and correlation between neighboring Q -points. This unified framework—combining basis projection with smooth priors—provides a physically grounded and computationally flexible approach to extract latent information from noisy, multidimensional scattering data.

4.3 Case III: X-ray Photon Correlation Spectroscopy (XPCS)

The physical quantity measured in X-ray Photon Correlation Spectroscopy (XPCS) is the time autocorrelation of the coherent scattering intensity at a fixed scattering vector Q . For simplicity, we fix Q and focus on the temporal signal $I(t)$, representing the time-dependent intensity at the detector. Due to the extremely limited photon counts in typical XPCS experiments, **direct fitting of $I(t)$ to a nonlinear model is generally not feasible**. Nonetheless, the time autocorrelation function $C(\tau)$ is expected to vary

smoothly with the lag time τ , providing a meaningful assumption that can be exploited through Bayesian inference.

Rather than attempting to reconstruct $C(\tau)$ directly from $I(t)$, we leverage the Wiener–Khinchin theorem, which relates the autocorrelation function to the power spectral density of the signal. This spectral-domain reformulation enables us to linearize the inference problem and naturally incorporate uncertainty propagation through Fourier transforms.

- **Prior:** The autocorrelation function $\mathbf{C} = [C(\tau_1), \dots, C(\tau_L)]^\top \in \mathbb{R}^L$ is assumed to be smooth and follows a multivariate Gaussian distribution:

$$p(\mathbf{C}) = \frac{1}{Z'} \exp \left(-\frac{1}{2} \mathbf{C}^\top \mathbf{K}^{-1} \mathbf{C} \right), \quad (49)$$

where the prior covariance matrix $\mathbf{K} \in \mathbb{R}^{L \times L}$ is defined as

$$K_{\ell\ell'} = \sigma^2 \exp \left(-\frac{(\tau_\ell - \tau_{\ell'})^2}{2\lambda^2} \right). \quad (50)$$

- **Likelihood:** The power spectral density $\mathbf{H}^{\text{Exp}} \in \mathbb{R}^N$ is computed from the noisy time-resolved intensity signal $\mathbf{I}^{\text{Exp}} \in \mathbb{R}^M$, measured at time points t_m . Define the discrete Fourier transform (DFT) matrix $\mathbf{F} \in \mathbb{C}^{N \times M}$ by

$$\mathbf{F}_{nm} = \exp(-i\omega_n t_m), \quad (51)$$

where ω_n are the angular frequencies of the sampled modes. The frequency-domain signal is

$$\tilde{\mathbf{I}} = \mathbf{F} \mathbf{I}^{\text{Exp}}, \quad H_n^{\text{Exp}} = |\tilde{I}_n|^2, \quad (52)$$

$$\mathbf{H}^{\text{Exp}} = (\mathbf{F} \mathbf{I}^{\text{Exp}}) \odot (\mathbf{F} \mathbf{I}^{\text{Exp}})^\dagger, \quad (53)$$

with $\tilde{\mathbf{I}} \in \mathbb{C}^N$ the Fourier-transformed signal and \odot denotes the element-wise (Hadamard) product.

The time-domain uncertainty is characterized by a diagonal covariance matrix $\Sigma_I = \text{diag}((\Delta I_1)^2, \dots, (\Delta I_M)^2)$. Propagating this through the DFT, and applying the variance formula for squared modulus of complex Gaussian variables, we approximate the spectral-domain uncertainty by a diagonal covariance matrix $\Sigma_H \in \mathbb{R}^{N \times N}$:

$$[\Sigma_H]_{nn} = 2 \cdot (\mathbf{F} \Sigma_I \mathbf{F}^\dagger)_{nn} \cdot |\tilde{I}_n|^2 + ((\mathbf{F} \Sigma_I \mathbf{F}^\dagger)_{nn})^2. \quad (54)$$

We assume uncorrelated measurement noise in time and negligible correlation across frequencies.

According to the Wiener–Khinchin theorem, the theoretical power spectral density $\mathbf{H} \in \mathbb{R}^N$ is the Fourier transform of the autocorrelation function $\mathbf{C} \in \mathbb{R}^L$. Considering \mathbf{C} is symmetric, the Direct cosine transform (DCT) matrix is applied: $\mathcal{F} \in \mathbb{C}^{N \times L}$:

$$\mathbf{H} = \mathcal{F}\mathbf{C}, \quad \mathcal{F}_{n\ell} = \cos(\omega_n \tau_\ell). \quad (55)$$

Assuming Gaussian noise in the spectral domain, the likelihood becomes:

$$p(\mathbf{H}^{\text{Exp}} | \mathbf{C}) = \frac{1}{Z} \exp \left(-\frac{1}{2} (\mathbf{H}^{\text{Exp}} - \mathcal{F}\mathbf{C})^\top \boldsymbol{\Sigma}_H^{-1} (\mathbf{H}^{\text{Exp}} - \mathcal{F}\mathbf{C}) \right). \quad (56)$$

- **Objective:** Applying Bayes' theorem, the posterior distribution is

$$p(\mathbf{C} | \mathbf{H}^{\text{Exp}}) \propto p(\mathbf{H}^{\text{Exp}} | \mathbf{C}) \cdot p(\mathbf{C}), \quad (57)$$

which yields the negative log-posterior:

$$\mathcal{L}(\mathbf{C}) = \frac{1}{2} (\mathbf{H}^{\text{Exp}} - \mathcal{F}\mathbf{C})^\top \boldsymbol{\Sigma}_H^{-1} (\mathbf{H}^{\text{Exp}} - \mathcal{F}\mathbf{C}) + \frac{1}{2} \mathbf{C}^\top \mathbf{K}^{-1} \mathbf{C}. \quad (58)$$

Following the procedure mentioned in previous sections, minimizing $J(\mathbf{C})$ yields the MAP estimate of the autocorrelation function:

- **Forward Form:**

$$\mathbf{C}^{\text{MAP}} = (\mathcal{F}^\top \boldsymbol{\Sigma}_H^{-1} \mathcal{F} + \mathbf{K}^{-1})^{-1} \mathcal{F}^\top \boldsymbol{\Sigma}_H^{-1} \mathbf{H}^{\text{Exp}}, \quad (59)$$

or equivalently,

- **Dual Form:**

$$\mathbf{C}^{\text{MAP}} = \mathbf{K} \mathcal{F}^\top (\mathcal{F} \mathbf{K} \mathcal{F}^\top + \boldsymbol{\Sigma}_H)^{-1} \mathbf{H}^{\text{Exp}}. \quad (60)$$

This framework enables indirect inference of time autocorrelations from low-flux, high-noise intensity measurements by incorporating spectral-domain uncertainty and leveraging priors for smoothness and regularization.

5 Inference under Symmetry Assumptions

In the analysis of two-dimensional (2D) small-angle scattering data, the scattering intensity is represented as a function of the in-plane scattering vector components, $I(Q_x, Q_y)$. Owing to the inherent rotational symmetries of many soft-matter and biological systems, this intensity is often more naturally expressed in polar coordinates (q, θ) , where $q = \sqrt{Q_x^2 + Q_y^2}$ is the magnitude of the scattering vector and $\theta = \text{atan2}(Q_y, Q_x)$ denotes the azimuthal angle.

To reconstruct $I(q, \theta)$ from sparse or noisy measurements, we adopt a Gaussian Process Regression (GPR) framework that combines radial smoothness in q with a localized angular basis expansion in θ . This formulation allows for accurate denoising and interpolation while preserving the physical symmetries of the underlying scattering signal.

We model the angular dependence of the scattering function using a set of basis functions that respect the symmetry $I(q, \theta) = I(q, \theta + \pi)$. To enforce this, we use Fourier series:

$$\phi_n(\theta) = \begin{cases} 1 & \text{if } n = 0 \\ \cos(2n(\theta - \theta_0)) & \text{for } n = 1, 2, \dots, N. \end{cases} \quad (61)$$

The center θ_0 is determined by the eigenvector of the gyration tensor of scattering intensity.

The angular expansion of the anisotropic component of scattering intensity is written as:

$$\tilde{I}(q, \theta) = \sum_{n=0}^N A_n(q) \phi_n(\theta), \quad (62)$$

where $A_0(q)$ captures the isotropic component and higher-order $A_n(q)$ terms encode anisotropic features.

$$\tilde{I}(q, \theta) = \frac{I(q, \theta)}{I^{\text{Iso}}(q)} \quad (63)$$

$$\Delta \tilde{I}(q, \theta) = \frac{\Delta I(q, \theta)}{I^{\text{Iso}}(q)} \quad (64)$$

and

$$I^{\text{Iso}}(q) = \frac{1}{2\pi} \int_0^{2\pi} I(q, \theta) d\theta \quad (65)$$

We define a row-normalized spatial weight matrix to map between observed and evaluation coordinates:

$$W_{ai} = \frac{\exp\left(-\frac{|q_a - q_i|^2}{2\mu^2}\right)}{\sum_{j=1}^L \exp\left(-\frac{|q_a - q_j|^2}{2\mu^2}\right)}, \quad (66)$$

where μ governs the locality of the spatial coupling.

We adopt the following indexing convention: the subscript a denotes the observed data points, i indexes the inference grid points along the radial coordinate q_i , and n labels the angular basis functions. Accordingly, the observed (possibly transformed) intensity is written as

$$\tilde{I}_a^{\text{Exp}} = \tilde{I}^{\text{Exp}}(q_a, \theta_a), \quad (67)$$

and the expansion coefficients at the inference points are denoted

$$A_{in} = A_n(q_i). \quad (68)$$

The expected intensity at each observation point is then modeled as:

$$\tilde{I}_a^{\text{Exp}} = W_{ai} \phi_n A_{in} + \eta_a, \quad (69)$$

where $\eta_a \sim \mathcal{N}(0, \Sigma_{ab})$ represents experimental noise with diagonal covariance:

$$\Sigma_{ab} = \delta_{ab} (\Delta \tilde{I}_a)^2. \quad (70)$$

To account for radial correlations in expansion coefficients A_{in} , we model the coefficients A_{in} as smooth functions of q_i using a radial basis function (RBF) kernel:

$$k_{ij} = \exp\left(-\frac{(q_i - q_j)^2}{2\lambda^2}\right), \quad (71)$$

where λ is a length-scale hyperparameter controlling radial smoothness. Moreover, as empirically the expansion coefficients decays as

$$|A_{in}| \sim n^{-2} \quad (72)$$

we define a diagonal angular penalty matrix

$$B_{nq} = n^2 \delta_{nq}, \quad (73)$$

The prior over the coefficients is given by:

$$A_{in} \sim \mathcal{N}(0, K_{in,jq}), \quad \text{with } K_{in,jq} = k_{ij} B_{nq}. \quad (74)$$

We perform Bayesian inference by minimizing the following objective function:

$$\mathcal{L}(A) = (I_a^{\text{Exp}} - G_{a,in} A_{in}) \Sigma_{ab}^{-1} (I_b^{\text{Exp}} - G_{b,jq} A_{jq}) + (A_{in} K_{in,jq}^{-1} A_{jq}), \quad (75)$$

where $G_{a,in} = W_{ai} \phi_n$ is the forward mapping.

The maximum *a posteriori* (MAP) estimation of the coefficients are then obtained via GPR:

$$A_{in}^{\text{MAP}} = K_{b,in}^G \tilde{\Sigma}_{ab}^{-1} I_b^{\text{Exp}}, \quad (76)$$

with

$$K_{a,in}^G = K_{in,jq} G_{a,jq}, \quad (77)$$

$$\tilde{\Sigma}_{ab} = \Sigma_{ab} + G_{a,in} K_{in,jq} G_{b,jq}. \quad (78)$$

The reconstructed scattering intensity at new coordinates (q_i, θ_i) is given by:

$$I^{\text{MAP}}(q_i, \theta_i) = I_i^{\text{Iso}} \tilde{I}_i^{\text{MAP}}, \quad (79)$$

where

$$\tilde{I}_i^{\text{MAP}} = \phi_{in} W_{ij} A_{jn}^{\text{MAP}}. \quad (80)$$

The isotropic component at each evaluation coordinate i can be estimated directly from observed data,

$$I_i^{\text{Iso}} = W_{ia} I_a^{\text{Exp}}. \quad (81)$$

6 Inference under Irreversible Transformations

Beyond improving robustness to noise, the probabilistic inference framework is also well-suited for handling **irreversible transformations**, where the relationship between the observed data and the latent physical quantity is non-invertible or ill-posed. In such scenarios, the forward mapping from the true signal to the measured quantity is often not one-to-one, and the inverse problem lacks a unique solution.

A representative example is the effect of **instrumental resolution smearing**, where the measured scattering intensity $I^{\text{Exp}}(Q)$ is related to the true intensity $I(Q)$ through a convolution with a resolution function $R(Q, Q')$:

$$I^{\text{Exp}}(Q) = \int I(Q') R(Q, Q') dQ'. \quad (82)$$

Here, $R(Q, Q')$ is typically modeled as a Gaussian kernel:

$$R(Q, Q') = \frac{1}{\sqrt{2\pi\Delta_Q(Q)^2}} \exp\left(-\frac{(Q - Q')^2}{2\Delta_Q(Q)^2}\right), \quad (83)$$

where $\Delta_Q(Q)$ is the resolution width, often dependent on the scattering vector Q . This convolution operation smooths or blurs high-frequency components in the true spectrum, making the inversion ill-posed, which means multiple candidate functions $I(Q)$ may produce the same observed profile.

To make the inversion feasible, we adopt a Bayesian regularization strategy. Specifically, we impose prior knowledge on $I(Q)$ and incorporate the measurement model into a probabilistic framework:

- **Prior:** A natural assumption is that $I(Q)$ varies smoothly with Q , which can be captured by a Gaussian process prior:

$$p(\mathbf{I}) = \frac{1}{Z'} \exp\left(-\frac{1}{2}\mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I}\right), \quad (84)$$

where $\mathbf{I} \in \mathbb{R}^L$ is the vector of intensities evaluated at discrete Q_i points, and $\mathbf{K}_{ij} = \sigma^2 \exp\left(-\frac{(Q_i - Q_j)^2}{2\lambda^2}\right)$ is the prior covariance matrix enforcing smoothness.

- **Likelihood:** Discretizing Eqn. (82), we express the forward model as:

$$\mathbf{I}^{\text{Exp}} = \mathbf{R}\mathbf{I} + \boldsymbol{\eta}, \quad (85)$$

where $\mathbf{I}^{\text{Exp}} \in \mathbb{R}^M$ is the observed intensity vector, and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ models measurement noise with known covariance $\boldsymbol{\Sigma}$. Each entry of the resolution matrix $\mathbf{R} \in \mathbb{R}^{M \times L}$ is defined by:

$$R_{ij} = \frac{\exp\left(-\frac{(Q_i - Q_j)^2}{2\Delta_Q(Q_i)^2}\right)}{\sum_{k=1}^L \exp\left(-\frac{(Q_i - Q_k)^2}{2\Delta_Q(Q_i)^2}\right)}, \quad (86)$$

where Q_i denotes the detector coordinate (observation location), Q_j is the latent coordinate where the true intensity is evaluated, and $\Delta_Q(Q_i)$ represents the instrument resolution width at point Q_i . This Gaussian kernel formulation ensures each row of \mathbf{R} is normalized to sum to unity:

$$\sum_{j=1}^L R_{ij} = 1 \quad \text{for all } i = 1, \dots, M, \quad (87)$$

thereby preserving total intensity under convolution. Therefore

$$p(\mathbf{I}^{\text{Exp}} | \mathbf{I}) = \frac{1}{Z} \exp\left(-\frac{1}{2}(\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})\right). \quad (88)$$

- **Objective:** By Bayes' theorem, the posterior over \mathbf{I} is proportional to the product of the likelihood and prior:

$$p(\mathbf{I} | \mathbf{I}^{\text{Exp}}) \propto \exp\left(-\frac{1}{2}(\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})\right) \cdot \exp\left(-\frac{1}{2}\mathbf{I}^\top \mathbf{K}^{-1}\mathbf{I}\right). \quad (89)$$

The negative log-posterior yields the MAP objective function:

$$\mathcal{L}(\mathbf{I}) = (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I}) + \mathbf{I}^\top \mathbf{K}^{-1}\mathbf{I}. \quad (90)$$

Minimizing $\mathcal{L}(\mathbf{I})$ yields the MAP estimate of the scattering intensity:

- **Forward Form:**

$$\mathbf{I}^{\text{MAP}} = (\mathbf{R}^\top \boldsymbol{\Sigma}^{-1} \mathbf{R} + \mathbf{K}^{-1})^{-1} \mathbf{R}^\top \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}, \quad (91)$$

- **Dual Form:**

$$\mathbf{I}^{\text{MAP}} = \mathbf{K} \mathbf{R}^\top (\mathbf{R} \mathbf{K} \mathbf{R}^\top + \boldsymbol{\Sigma})^{-1} \mathbf{I}^{\text{Exp}}, \quad (92)$$

Alternatively, one can also apply the smooth prior indirectly to the scattering intensity by regularizing the underlying representation. Consider first that \mathbf{I} can be expressed as a basis expansion in the form:

$$\mathbf{I} = \mathbf{C}\mathbf{A}, \quad (93)$$

where $\mathbf{C} \in \mathbb{R}^{L \times N}$ is a matrix of basis functions evaluated at points Q_i , and $\mathbf{A} \in \mathbb{R}^N$ is the corresponding coefficient vector. The effect of instrumental resolution smearing on the expanded intensity becomes:

$$\mathbf{I}^{\text{Exp}} = \mathbf{R}\mathbf{C}\mathbf{A} + \boldsymbol{\eta}, \quad (94)$$

where $\mathbf{R} \in \mathbb{R}^{M \times L}$ is the resolution matrix, and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is Gaussian noise.

To streamline notation, we define the combined forward operator:

$$\mathbf{G} = \mathbf{R}\mathbf{C}, \quad (95)$$

so that the observation model becomes:

$$\mathbf{I}^{\text{Exp}} = \mathbf{G}\mathbf{A} + \boldsymbol{\eta}. \quad (96)$$

Assuming the coefficients \mathbf{A} vary smoothly with respect to Q , we place a Gaussian prior:

$$p(\mathbf{A}) = \frac{1}{Z'} \exp \left(-\frac{1}{2} \mathbf{A}^\top \mathbf{K}^{-1} \mathbf{A} \right), \quad (97)$$

where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the prior covariance matrix that enforces smoothness in the coefficient space.

The MAP estimate of \mathbf{A} is obtained by minimizing the negative log-posterior, resulting in the following closed-form solutions:

$$\mathbf{A}^{\text{MAP}} = (\mathbf{G}^\top \boldsymbol{\Sigma}^{-1} \mathbf{G} + \mathbf{K}^{-1})^{-1} \mathbf{G}^\top \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}, \quad (98)$$

$$\mathbf{A}^{\text{MAP}} = \mathbf{K} \mathbf{G}^\top (\mathbf{G} \mathbf{K} \mathbf{G}^\top + \boldsymbol{\Sigma})^{-1} \mathbf{I}^{\text{Exp}}. \quad (99)$$

7 Inference under Chromatic Aberrations

In elastic scattering experiments, the detector typically records an intensity that is averaged over a range of incident wavelengths. Traditional instrument design addresses this chromatic dispersion by narrowing the wavelength distribution using energy-selective components, such as rotating choppers or monochromator crystals. However, these selection mechanisms significantly reduce the usable neutron flux by discarding a large fraction of the source spectrum. To better utilize the full intensity of the neutron source, it is therefore important to develop correction algorithms that can account for chromatic aberrations during data analysis. In this section, we derive the transformation function that relates the intrinsic scattering intensity to the detector-recorded signal under chromatic aberration. This relationship is expressed in the form of a transformation matrix, which is then integrated into the probabilistic inference framework.

Since the scattering angle θ depends on both the wavelength λ and the momentum transfer q , this averaging introduces a spatial rescaling of the observed scattering pattern. We derive a forward model that expresses the measured intensity as a linear transformation of a latent intensity profile defined at a reference wavelength λ_0 .

The momentum transfer is given by

$$q(\theta, \lambda) = \frac{4\pi}{\lambda} \sin\left(\frac{\theta}{2}\right), \quad (100)$$

where the scattering angle θ is determined from the detector displacement from beam center d , and sample-to-detector distance L via

$$\theta(d) = \arctan\left(\frac{d}{L}\right). \quad (101)$$

Combining these expressions, the wavelength-dependent momentum transfer at position d is

$$q(d, \lambda) = \frac{4\pi}{\lambda} \sin\left(\frac{1}{2} \arctan\left(\frac{d}{L}\right)\right). \quad (102)$$

We assume that for each incident wavelength λ_i , the intensity observed at position d_k corresponds to the intensity of a latent profile evaluated at a rescaled coordinate d'_{ik} under a fixed reference wavelength λ_0 , satisfying

$$I(d_k, \lambda_i) = I(d'_{ik}, \lambda_0), \quad \text{where} \quad d'_{ik} = f_d(d_k, \lambda_i) = q^{-1}(q(d_k, \lambda_i), \lambda_0). \quad (103)$$

Given a normalized wavelength distribution $w(\lambda)$, the measured intensity at detector displacement d_k is the weighted average of all contributing wavelengths:

$$I^{\text{Exp}}(d_k) = \sum_{i=1}^N w(\lambda_i) \cdot I(d_k, \lambda_i) = \sum_{i=1}^N w(\lambda_i) \cdot I(d'_{ik}, \lambda_0). \quad (104)$$

To evaluate $I(d'_{ik}, \lambda_0)$ on a fixed detector grid $\{d_j\}_{j=1}^M$, we apply kernel interpolation using a radial basis function $\phi(d_j, d'_{ik})$, defined as:

$$\phi(d_j, d'_{ik}) = \frac{\exp\left(-\frac{(d_j - d'_{ik})^2}{2\mu^2}\right)}{\sum_{j'=1}^M \exp\left(-\frac{(d_{j'} - d'_{ik})^2}{2\mu^2}\right)}, \quad (105)$$

where μ is a kernel width parameter. This yields the approximation:

$$I(d'_{ik}, \lambda_0) = \sum_{j=1}^M \phi(d_j, d'_{ik}) \cdot I(d_j, \lambda_0). \quad (106)$$

Substituting into Eq. (104), the full expression for the measured intensity becomes

$$I^{\text{Exp}}(d_k) = \sum_{i=1}^N \sum_{j=1}^M w(\lambda_i) \cdot \phi(d_j, d'_{ik}) \cdot I(d_j, \lambda_0). \quad (107)$$

We now express this operation in matrix form. Let $\mathbf{I}^{\text{Exp}} \in \mathbb{R}^M$ denote the observed intensity vector with entries $I^{\text{Exp}}(d_k)$, and $\mathbf{I} \in \mathbb{R}^M$ the latent profile at the reference wavelength λ_0 . Define the resampling matrix $\mathbf{R} \in \mathbb{R}^{M \times M}$ by

$$R_{kj} = \sum_{i=1}^N w(\lambda_i) \cdot \phi(d_j, f_d(d_k, \lambda_i)), \quad (108)$$

so that the forward model becomes

$$\mathbf{I}^{\text{Exp}} = \mathbf{R}\mathbf{I}. \quad (109)$$

All entries in the matrix \mathbf{R} are explicitly determined by known parameters: the detector geometry $\{d_j\}$, the sample-to-detector distance L , the interpolation kernel width μ , and the wavelength distribution $w(\lambda)$. Thus,

the averaging process is a deterministic, wavelength-encoded resampling of the latent profile.

This formulation establishes a clear forward transform model for probabilistic inference to recover the monochromatic scattering profile. Because \mathbf{R} is independent of the unknown intensity \mathbf{I} , the measurement \mathbf{I}^{Exp} can be interpreted as a known linear transformation of \mathbf{I} .

- **Prior:** We posit that the latent intensity profile $I(d)$ is a smooth function of detector position. This assumption is implemented via a Gaussian process prior:

$$p(\mathbf{I}) = \frac{1}{Z'} \exp \left(-\frac{1}{2} \mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I} \right), \quad (110)$$

where $\mathbf{K} \in \mathbb{R}^{M \times M}$ is a covariance matrix defined by a kernel such as

$$\mathbf{K}_{jk} = \sigma^2 \exp \left(-\frac{(d_j - d_k)^2}{2\lambda^2} \right), \quad (111)$$

with hyperparameters σ and λ controlling the amplitude and correlation length of intensity variations.

- **Likelihood:** Assuming Gaussian measurement noise, the observed intensity is modeled as:

$$\mathbf{I}^{\text{Exp}} = \mathbf{R}\mathbf{I} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (112)$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ encodes the experimental uncertainty. The likelihood function becomes:

$$p(\mathbf{I}^{\text{Exp}} | \mathbf{I}) = \frac{1}{Z} \exp \left(-\frac{1}{2} (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I}) \right). \quad (113)$$

- **Posterior and Objective:** By Bayes' theorem, the posterior over \mathbf{I} is proportional to the product of likelihood and prior:

$$p(\mathbf{I} | \mathbf{I}^{\text{Exp}}) \propto \exp \left(-\frac{1}{2} (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I}) \right) \exp \left(-\frac{1}{2} \mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I} \right). \quad (114)$$

Taking the negative log of the posterior yields the MAP objective function:

$$\mathcal{L}(\mathbf{I}) = (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{I}^{\text{Exp}} - \mathbf{R}\mathbf{I}) + \mathbf{I}^\top \mathbf{K}^{-1} \mathbf{I}. \quad (115)$$

Minimizing this loss gives the MAP estimate of the latent profile:

- **Forward Form:**

$$\mathbf{I}^{\text{MAP}} = (\mathbf{R}^\top \boldsymbol{\Sigma}^{-1} \mathbf{R} + \mathbf{K}^{-1})^{-1} \mathbf{R}^\top \boldsymbol{\Sigma}^{-1} \mathbf{I}^{\text{Exp}}, \quad (116)$$

- **Dual Form:**

$$\mathbf{I}^{\text{MAP}} = \mathbf{K} \mathbf{R}^\top (\mathbf{R} \mathbf{K} \mathbf{R}^\top + \boldsymbol{\Sigma})^{-1} \mathbf{I}^{\text{Exp}}. \quad (117)$$

The MAP estimate \mathbf{I}^{MAP} can now be computed from the measured intensity \mathbf{I}^{Exp} , the associated uncertainty covariance $\boldsymbol{\Sigma}$, and the known wavelength distribution $w(\lambda)$.

8 Inference under Extremely Low Detector Count

In scattering experiments with weakly scattering samples or limited exposure times, the number of detected scattering events per detector bin is often small. In such low-count regimes, the assumption of additive Gaussian noise becomes invalid, and the measurement noise is better modeled by a **Poisson distribution**, which captures the discrete, non-negative nature of the observed counts.

Let the measured data be $\mathbf{y} = [y_1, y_2, \dots, y_M]^\top$, where each $y_i \in \mathbb{N}$ represents the number of scattering events at detector position Q_i . The expected number of counts is modeled as:

$$\boldsymbol{\mu} = \mathbf{GA}, \quad (118)$$

where $\mathbf{G} \in \mathbb{R}^{M \times N}$ is the forward operator (encoding instrumental resolution and basis representation), and $\mathbf{A} \in \mathbb{R}^N$ is the latent coefficient vector to be inferred.

- **Prior:** A Gaussian prior is placed on \mathbf{A} , promoting smoothness via a positive-definite covariance matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$:

$$p(\mathbf{A}) = \frac{1}{Z'} \exp \left(-\frac{1}{2} \mathbf{A}^\top \mathbf{K}^{-1} \mathbf{A} \right). \quad (119)$$

- **Likelihood:** Each observation y_i is modeled as an independent Poisson random variable with mean $\mu_i = [\mathbf{GA}]_i$:

$$p(y_i | \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \quad (120)$$

$$p(\mathbf{y} | \mathbf{A}) = \prod_{i=1}^M \frac{[\mathbf{GA}]_i^{y_i} e^{-[\mathbf{GA}]_i}}{y_i!}. \quad (121)$$

- **MAP Objective:** Maximizing the posterior $p(\mathbf{A} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{A})p(\mathbf{A})$ is equivalent to minimizing the negative log-posterior:

$$\mathcal{L}(\mathbf{A}) = \sum_{i=1}^M ([\mathbf{GA}]_i - y_i \log([\mathbf{GA}]_i)) + \frac{1}{2} \mathbf{A}^\top \mathbf{K}^{-1} \mathbf{A} + \text{const}, \quad (122)$$

where the constant includes terms like $\log(y_i!)$ that do not depend on \mathbf{A} .

8.1 Minimization Algorithm

Since the gradient of the MAP objective can be derived explicitly:

$$\nabla_{\mathbf{A}} \mathcal{L}(\mathbf{A}) = \mathbf{G}^\top \left(\mathbf{1} - \frac{\mathbf{y}}{\mathbf{GA}} \right) + \mathbf{K}^{-1} \mathbf{A}, \quad (123)$$

gradient-based iterative minimization algorithms, such as gradient descent or quasi-Newton methods, can be used to find the optimal \mathbf{A} .

A suitable initial guess can be obtained from the Gaussian approximation:

$$\mathbf{A}^{(1)} = (\mathbf{G}^\top \Sigma^{-1} \mathbf{G} + \mathbf{K}^{-1})^{-1} \mathbf{G}^\top \Sigma^{-1} \mathbf{y}, \quad (124)$$

$$\Sigma = \text{diag}(\mathbf{y} + \epsilon), \quad (125)$$

where $\epsilon > 0$ is a small constant added to avoid singularities when $y_i = 0$.

We then proceed iteratively using **gradient descent** to minimize the MAP objective. At each iteration $j = 1, 2, \dots$, we update \mathbf{A} according to:

$$\mathbf{A}^{(j+1)} = \mathbf{A}^{(j)} - \alpha_j \nabla_{\mathbf{A}} \mathcal{L}(\mathbf{A}^{(j)}), \quad (126)$$

where $\alpha_j > 0$ is the step size.

Alternatively, a change of variable can be applied to prevent invalid evaluations of the Poisson log-likelihood when the predicted intensities $\boldsymbol{\mu} = \mathbf{GA}$ contain non-positive values. Specifically, we introduce an unconstrained variable $\mathbf{u} \in \mathbb{R}^M$ and define:

$$\mathbf{GA} = \exp(\mathbf{u}), \quad (127)$$

where the exponential is applied element-wise. This reparameterization ensures that the expected counts $\mu_i = [\mathbf{GA}]_i$ remain strictly positive for all i , as required by the Poisson likelihood.

Given $\mathbf{GA} = \exp(\mathbf{u})$, a consistent solution for \mathbf{A} is obtained:

$$\mathbf{A}(\mathbf{u}) = \mathbf{G}^\dagger \exp(\mathbf{u}). \quad (128)$$

where $\mathbf{G}^\dagger = (\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top$ denotes the Moore–Penrose pseudoinverse of \mathbf{G} . By the chain rule, the gradient with respect to \mathbf{u} is given by

$$\nabla_{\mathbf{u}} \mathcal{L}(\mathbf{u}) = \exp(\mathbf{u}) - \mathbf{y} + \text{diag}(\exp(\mathbf{u})) \cdot (\mathbf{G}^\dagger)^\top \mathbf{K}^{-1} \mathbf{G}^\dagger \exp(\mathbf{u}) \quad (129)$$

Gradient-based optimization can then proceed on the variable \mathbf{u} .

Experimental Technique	Observable	Likelihood	Inference Target	Prior
SANS	$I^{\text{Exp}}(Q) \pm \Delta I$	$\mathcal{N}(I^{\text{Exp}} I, \Sigma)$	Scattering profile $I(Q)$	Smooth in Q
NSE	$S^{\text{Exp}}(Q, t) \pm \Delta S$	$\mathcal{N}(S^{\text{Exp}} GA, \Sigma)$	Prony coefficients $A(Q, \tau)$	Smooth in Q
CVSANS	$I^{\text{Exp}}(Q, \gamma) \pm \Delta I$	$\mathcal{N}(I^{\text{Exp}} MS, \Sigma)$	Partial structure factors $S_n(Q)$	Smooth in Q
Desmearing	$I^{\text{Exp}}(Q) \pm \Delta I$	$\mathcal{N}(I^{\text{Exp}} RI, \Sigma)$	Resolution-corrected profile $I(Q)$	Smooth in Q
2D Scattering	$I^{\text{Exp}}(Q, \theta) \pm \Delta I$	$\mathcal{N}(I^{\text{Exp}} \Phi A, \Sigma)$	Fourier series coefficients $A_n(Q)$	Smooth in Q
XPCS	$I^{\text{Exp}}(t) \pm \Delta I$	$\mathcal{N}(H^{\text{Exp}} \mathcal{FC}, \Sigma_H)$	Autocorrelation $C(\tau)$	Smooth in τ

Table 1: Summary of probabilistic inference formulations for various scattering techniques, highlighting the observable quantities, statistical likelihoods, inferred physical parameters, and prior assumptions. Measurement uncertainties are explicitly incorporated in the observed data.

References

- [1] Claude E Shannon. A mathematical theory of communication. *BSTJ*, 27(3):379–423, 1948.
- [2] Harold Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, UK, 3rd edition, 1961.
- [3] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. CRC Press, Boca Raton, 3rd edition, 2013.
- [4] Christopher K. I. Williams and Carl Edward Rasmussen. Gaussian processes for regression. *Adv. Neural Inf. Process. Syst.*, 8:514–520, 1996.
- [5] Christopher K. I. Williams and David Barber. Bayesian classification with gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(12):1342–1351, 1998.
- [6] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006.
- [7] Nicholas W. Tschoegl. *The Phenomenological Theory of Linear Viscoelastic Behavior: An Introduction*. Springer-Verlag, Berlin, 1989.

- [8] Yves Nievergelt. *Prony's Method: A Mathematical Perspective*. Springer, Boston, 2000.
- [9] Victor Pereyra and Godela Scherer. *Exponential Data Fitting and Its Applications*. Bentham Science, Sharjah, 2010.
- [10] John C Mauro and Yihong Z Mauro. On the prony series representation of stretched exponential relaxation. *Physica A: Statistical Mechanics and its Applications*, 506:75–87, 2018.
- [11] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.
- [12] James G. Simmonds. *A Brief on Tensor Analysis*. Springer, New York, 1994.
- [13] Koichi Mayumi, Tatsuro Oda, Shinya Miyajima, Ippei Obayashi, and Kazuaki Tanaka. Error evaluation of partial scattering functions obtained from contrast-variation small-angle neutron scattering. *J. Appl. Cryst.*, 58:4–17, 2025.