To appear in Journal of the Optical Society of America A LIDS-P-1992

Bayesian Signal Reconstruction, Markov Random Fields, and X-Ray Crystallography

Peter C. Doerschuk

School of Electrical Engineering

Purdue University

West Lafayette, Indiana 47907–1285

(317) 494–1742

February 24, 1991

Abstract

A signal reconstruction problem motivated by X-ray crystallography is (approximately) solved in a Bayesian statistical approach. The signal is zero-one, periodic, and substantial statistical a priori information is known, which is modeled with a Markov random field. The data are inaccurate magnitudes of the Fourier coefficients of the signal. The solution is explicit and the computational burden is independent of the signal dimension. The spherical model and asymptotic small-noise expansions are used.

1 Introduction

I present a novel Bayesian statistical approach to a class of signal-reconstruction problems exemplified by the inverse problem of single-crystal X-ray crystallography. The signal is reconstructed from a substantial amount of a priori information plus inaccurate measurements of the magnitude of its Fourier transform. The major qualitative difficulty, even more prominent than in standard phase-retrieval problems, is the need to simultaneously treat constraints in both Object and Fourier space.

In more detail, the problem formulation and solution have the following properties:

- The available data are inaccurate observations of the magnitudes of the Fourier coefficients of the signal.
- The signal is zero-one.
- Substantial a priori statistical information concerning the pattern of zeros and ones is available.
- The signal is periodic (or is invariant under the action of a more general space group).
- The approach is Bayesian.
- The a priori knowledge in the Bayesian formulation is stated in the form of the probability density of a finite-lattice Markov random field (MRF) with a shift-invariant but otherwise arbitrary quadratic Hamiltonian.
- The use of the MRF prior allows the a priori atomic locations to be correlated in the crystallography application.
- The a posteriori probability density is also a MRF.
- The solution is explicit, i.e., no numerical quadratures or nonlinear programming problems are required.
- The signal is discretized and the computational burden is proportional to the number of samples and is independent of the dimension (i.e., 1D time signal, 2D image, etc.).
- The solution allows missing data and data samples with different levels of observation

noise.

• The hypercube constraint of the zero-one signal is approximated by a hypersphere constraint ("Spherical Model").

• The estimator is computed asymptotically as the observation noise tends to zero.

This paper is motivated by an inverse problem for the simplest physical model of an X-ray crystallography experiment. At the X-ray wavelengths of interest (1-2 Angstrom), the interaction of the radiation with the crystal is primarily elastic scattering from the electron density. The electron density is modeled as a periodic collection of identical impulses in d-space, one impulse for each atom in the crystal. The assumption that the impulses are identical is quite reasonable for organic molecules where the important atoms are C, N, and O. Note that the calculations are essentially independent of d, the dimension of the space. Because of the geometry of the usual experimental arrangement and the fact that the interaction of the radiation and the crystal is weak, the scattering is the Fourier transform of the scatterer, in this case the electron density. Because of limitations in detector technology, only the magnitude and not the phase of the scattering can be recorded. This magnitude function, called the diffraction pattern, is the fundamental experimental data.

The goal of the inverse problem is to compute the position of each atom in the molecule(s) making up a unit cell[1] of the crystal given imprecise measurements of a diffraction pattern from the crystal and some amount of a priori information concerning the nature of the scatterer. The fundamental difficulty in this inverse problem is that the measurements are related to the Fourier transform of the scatterer while the a priori knowledge is related to the scatterer itself.

The a priori information is a schematic summary of some knowledge of chemistry. Various

[1]Crystal symmetry is described by the invariance of the crystal structure under the action of a space group. A crystal is constructed from a unit cell that is repeated by multiple translation along the three unit cell vectors. The asymmetric unit is a subset of the unit cell, reflecting symmetries within the unit cell, such that a function defined over the unit cell can be uniquely specified by its values over the asymmetric unit.

amounts of a priori information form different independently-interesting inverse problems. The simplest information is simply knowledge that the electron density is always positive. More detail is provided by including the atomicity of the electron density. In most experiments, the empirical formula of the molecule[2] making up the crystal or even the graph of covalent atomic bondings is known. An abbreviated form of the graph information is simply to know the range of valences for each type of atom. The previous information was basically deterministic in nature. There is also basically statistical information concerning typical bond lengths, bond angles, and atomic valences. A major theme of this work is to balance the detail of the a priori information with the complexity of the calculations required to exploit it.

Inverse problems of this type tailored to crystallography have been of major interest for half a century [1, 2]. Methods exist to routinely solve small molecules. For reasons discussed later in this section, medium ($\approx 10^2$ atoms per asymmetric unit) and large molecules are either much more difficult or unsolvable (or, for quite large molecules such as proteins, require different methods based on multiple diffraction patterns from specially chosen chemical derivatives of the molecule of interest [3]). Especially with the continued development of molecular biology techniques, the number of medium and large molecules whose geometrical structures are desired is steadily increasing. Therefore further development in inverse problem methods seems very desirable.

The most powerful existing methods for small molecules are probabilistic in nature [1, 2, 4, 5, 6, 7]. The methods for crystallography are compared and contrasted with the methods for imaging problems by Millane [8]. One important difference is that the periodic nature

[2]A crystal of a large biological molecule is roughly half (by volume) the molecule of interest and half solvent and ions. Selected solvent molecules and ions will be ordered and therefore appear in the diffraction pattern and the crystallographic structure. Furthermore, certain pieces of the biological molecule may be disordered and therefore not appear in the crystallographic structure. Therefore the list of atoms present in the unit cell of the crystallographic structure depends on more factors than just the empirical formula of the biological molecule of interest.

of the electron density leads to a forced undersampling of its Fourier transform magnitude. Among other sources, the book edited by Stark [9] describes a variety of approaches for phase-retrieval in imaging problems. Existing methods for small molecules view the problem as a phase retrieval problem. That is, they attempt to combine the inaccurate measured magnitude of the scattering with some amount of a priori information in order to compute an estimate of the unmeasured phase of the scattering. Then, with both magnitude and phase of the scattering, they compute an inverse Fourier transform which gives an estimate of the electron density function. Next, they locate the peaks of the electron density function and position atoms at those locations. (In actual practice, this is often an iterative process between Fourier and Coordinate spaces, and requires skilled human intervention). Finally, a weighted least squares optimization of the locations (and other parameters such as atomic vibrational temperatures) is performed. The weights are often derived from sample standard deviations of the measurements that are recorded during the course of the experiment.

The first difference between my approach and traditional methods is that I attempt to directly estimate the atomic locations without passing through an intermediate step of estimating scattering phase variables. There are two reasons for taking this approach. In many experiments there are many more scattering phases than atomic locations and therefore from a statistical point of view it is undesirable to first estimate the scattering phases. In addition, most good a priori models of atomic locations are in terms of positions rather than scattering phases.

Second, traditional methods use very simple models of atomic locations. They assume that the electron distribution is impulsive but that the locations of the impulses are independent identically (often uniformly) distributed random variables. A major component of my approach is to invest a great deal of effort in modeling of the correlations between the atomic locations. That is, I attempt to greatly improve the accuracy of the chemistry model. In the work described in this paper, these correlations are modeled in a purely statistical sense.

Third, traditional methods take a complicated view of the inaccuracies in the actual observations. These inaccuracies are due to photon counting statistics, detector errors, and deviations of the actual physical process from the idealized mathematical model. In current methods these inaccuracies are ignored at the phase-retrieval level, but included in the least squares optimization. My approach includes these inaccuracies in a fundamental way from the very start of the calculation.

The failure of current small molecule techniques to extend to larger molecules is attributed by Bricogne [6, 7] to

1. inconsistent usage of probabilistic information by the reconstruction of joint probability densities from marginal probability densities without accounting for the fact that certain data entered into multiple marginal distributions [6, Section 2.2.2] and

2. inaccurate computation of marginal probabilities by the use of Edgeworth expansions which are evaluated in the tails of the corresponding Gaussian distribution [6, Section 2.2.1].

Bricogne addresses (2) by computing multiple expansions centered at different trial positions and avoids (1) entirely by directly computing approximations to joint probability densities. The multiple expansion points are examined through a branching strategy. These ideas are closely related to maximum entropy through the introduction of independent but nonuniform a priori densities on the atomic locations. Impressed by the very idealized nature of the independent atomic location hypothesis[3] I take a different approach starting with a model where the atomic locations are not independent. Applying Bayesian ideas to this alternative model requires approximations, but the approximations appear to avoid the problems of (1) and (2) above.

Reflecting the differences between my approach and traditional methods, I use a rather different mathematical formulation and set of mathematical tools. As is standard in Bayesian

[3]This is also noted by Bricogne–see the final paragraph of [6, Section 1.1.1].

approaches, the statistical model separates into three parts, the a priori model which gives a probability measure on a collection of underlying random variables whose values are to be estimated, a transformation from the underlying random variables to the observed random variables, and the observational model which gives a conditional probability measure on the measured values of the observed random variables given their true values. The a priori model is a Markov random field (MRF), or equivalently in physics nomenclature a statistical lattice field theory, and the transformation and the observational model can also be integrated into this framework. As discussed above, the MRF allows for dependence between the different atomic locations. From the point of view of MRFs, this work is unusual because the Hamiltonian depends both on the field and on its Fourier transform. In addition, for many statistical estimation applications, the obvious Hamiltonian is not useful because it is invariant under translation, rotation, and reflection and therefore a symmetry breaking term must be added.

Motivated by [10], I introduce the spherical model to approximately deal with the mathematical difficulties due to the binary nature of the lattice variables. However, my mathematical treatment is very different. Specifically, [10, 11] use the scalar constraint of the spherical model as a δ-function weighting function, represent the δ-function through its Fourier transform, and make a nontrivial exchange of integration order before preceding to the large lattice limit. I, on the other hand, treat the constraint of the spherical model as the definition of a manifold and perform a Laplace type asymptotic evaluation of the multivariable integration over this manifold where the asymptotics is due to the observation error variances and where the critical point location is determined by the methods of constrained multivariable optimization theory. Independent of the spherical model, use of asymptotics in the variances of the observation errors rather than in the lattice spacing/number of lattice sites or what in a physics problem would be the external field strengths is unusual for lattice field theory calculations. It is these mathematical methods–the lattice field theory, symmetry breaking, spherical model, and small observation-error-variance asymptotics–that I wish to focus on

in this paper.

From a signal-processing point of view, it is important to note several aspects of the problem. First, the quality of the data varies greatly over the different observations. Therefore, estimation algorithms that can deal with varying observation noise are important. Second, some data points will not be present. Specifically, the low resolution data within a sphere centered around the DC Fourier coefficient and the high resolution data outside of an ellipsoid centered around the DC Fourier coefficient are absent. Therefore, estimation algorithms must also deal with missing data. Third, in this paper algorithms are proposed that provide estimates of atomic locations (and therefore phases) based on the experimental data. However, I do not claim that these algorithms have extracted all available information in the data. Rather I expect the results of these methods to be used as initial conditions for more computationally intensive nonlinear optimization algorithms, analogous to the nonlinear least squares used for refinement in crystallography.

An important part of crystallography is the space group symmetries of the crystal. In this paper only the most simple space group is considered, that is, the space group where the unit cell has no internal symmetries and therefore the asymmetric unit equals the unit cell. This space group is called P1. Furthermore, since the equations are essentially independent of d, the notation will be simplified by writing equations for $d = 1$ only. In the case $d = 1$ and P1, the only space group information is the single dimension of the unit cell.

The remainder of the paper is organized in the following fashion. The statistical model is presented in Section 2. In Section 3 the Bayesian statistical viewpoint, cost functions for a Bayesian estimator, and the need for an additional symmetry-breaking term in the Hamiltonian are discussed. Having presented the final Hamiltonian, the remainder of the calculation is outlined in Section 4. The spherical model is recalled in Section 5. After a change to Fourier coordinates (Section 6) and evaluation of angular integrals (Section 7), certain magnitude integrals are required which cannot be computed exactly. To treat this problem, asymptotics in the variances of the observation errors is introduced in Section 8.

Some notation and results of elementary calculations are collected together in Section 9. The critical point is computed in Section 10 using standard tools from constrainted multivariable optimization theory. The required asymptotic formulae are computed in Section 11. In Section 12 the previous results are combined to give formulae for the conditional expectation of the field. Finally, the results to date and directions for future research are discussed in Section 13.

2 Statistical Model

As described in the Introduction, in this paper a Bayesian view of the signal reconstruction problem is presented. The statistical model has three parts–the a priori model, the transformation from underlying to observed variables, and the observational model.

The physical model is that the electron density is made up of an infinite periodic collection of identical impulses normalized to unit amplitude, that the scattering is the Fourier transform of the electron density, and that the measured quantity is the magnitude squared of the scattering. The a priori probability measure describes how these impulses are positioned in space. The deterministic transformation is the Fourier transform followed by the magnitude-squared operation and thus comes directly from the physics of the problem. The conditional observational probability measure describes the errors in measuring the squared magnitudes and, in practice, also the inaccuracies of the physical model. Note that the a priori probability distribution is a much more subtle and flexible tool than merely specifying that the variables must belong to some function space, which would correspond to a probability measure that took on only two values–0 if the variables did not belong and an appropriate nonzero value if they did belong.

Place a lattice within the asymmetric unit of the unit cell and constrain the atoms to occupy sites in this lattice. This lattice is to be viewed as a numerical analysis lattice. The underlying random variables, denoted ϕ_n, are then taken as binary random variables, one at

each lattice site, where 0 (1) corresponds to absence (presence) of a generic atom at that site. The a priori probability measure is a Markov random field (MRF) [12] on this finite lattice. The desirable features of the MRF are that it can describe dependencies (corresponding to chemical bonds) between the atomic locations while at the same time it is sufficiently simple mathematically that calculations can be performed.

To describe the MRF it is necessary to specify a neighborhood structure (described at the end of this section), a set of boundary conditions, and an energy function (denoted H^{apriori}). The boundary conditions vary from space group to space group and for a given space group are the generalizations of toroidal boundary conditions that are implied by the space group. The energy function is the most general shift-invariant quadratic function, specifically,

$$\begin{aligned} u_n &= \sum_{n_1=0}^{L-1} \sum_{n_2=0}^{L-1} \phi_{n+n_1} w_2(n_1, n_2) \phi_{n+n_2} + w_1 \phi_n \\ H^{\text{apriori}} &= \sum_{n=0}^{L-1} u_n \end{aligned}$$

where, as discussed previously, only the case $d = 1$ with space group P1 is considered so it is necessary only to specify the periodicity of the crystal, which is L lattice sites. Without loss of generality it is possible to assume $w_2(n_1, n_2) = w_2(n_2, n_1)$ and to take the indicated form for the linear term and have no constant term. The a priori probability measure is then

$$\Pr(\{\phi\}) = \frac{1}{Z^{\text{apriori}}} e^{-H^{\text{apriori}}(\{\phi\})}$$

where Z^{apriori} is a normalization constant.

The reasons for this choice of H^{apriori} are two fold. First, this energy has a simple form in terms of the Fourier coefficients of the field. Specifically, it is a linear combination of functions of individual Fourier coefficients. Such simplicity is important for the success of these analytic calculations. Second, this form for the energy function contains a number of interesting special cases, including the Ising model of (anti) ferromagnetic materials and models that reasonably capture the bond-length limitations in covalently-bound molecules (see [13]). Much more detailed Hamiltonians, to which this quadratic Hamiltonian can be

viewed as an approximation, have been considered but such Hamiltonians are too complex for analytic calculations.

Denote the measured random variables as z_k, the measured values (which are inaccurate) as y_k, and the sample variances of the errors in the measured values as σ_k^2. Given the definition and interpretation of the ϕ_n and the simple physical model discussed previously, the deterministic transformation from underlying to measured variables is simply

$$\begin{aligned} z_k &= |\Phi_k|^2 \\ \Phi_k &= \sum_{n=0}^{L-1} \phi_n e^{-jnk\frac{2\pi}{L}}. \end{aligned}$$

Because a sample variance is measured for each reflection k (but no crosscorrelation information is measured) and because current methods use weighted least squares optimization, I have used a Gaussian assumption for the conditional observational probability measure. It is important to realize that this can also be written in the MRF formalism. Specifically, define

$$H^{\text{obs}} = \sum_{k=0}^{L-1} \frac{1}{2\sigma_k^2} (y_k - z_k(\{\phi\}))^2$$

where y_k and σ_k are observed in the experiment but I assume that σ_k is known exactly. Then

$$\Pr(\{y\}|\{\phi\}) = \frac{1}{Z^{\text{obs}}} e^{-H^{\text{obs}}(\{y\},\{\phi\})}.$$

The joint probability measure is

$$\Pr(\{y\},\{\phi\}) = \frac{1}{Z'} e^{-H'(\{y\},\{\phi\})}$$

where $H' = H^{\text{apriori}} + H^{\text{obs}}$ and $Z' = Z^{\text{apriori}} Z^{\text{obs}}$, and primes are used because the bulk of this paper will concern a modified Hamiltonian denoted H. Finally, the conditional observational probability measure conditional on the data is

$$\Pr(\{\phi\}|\{y\}) = \frac{\Pr(\{y\},\{\phi\})}{\Pr(\{y\})}.$$

For fixed $\{y\}$ this measure is proportional to the joint measure.

The a posteriori measure is also a MRF on the lattice variables $\{\phi\}$ with the same boundary conditions but with a different energy function (H') and a different neighborhood structure. The neighborhood structure for the a priori measure was determined essentially by the support of $w_2(\cdot,\cdot)$, could therefore have small neighborhoods (equivalently short range interactions), and therefore might allow efficient approximate computations based on disjoint neighborhood ideas. On the other hand, the neighborhood structure for the a posteriori measure is determined by the summation in the definition of the Fourier coefficients Φ_k, which makes every site a neighbor of every other site. This is one manifestation of the fundamental difficulty in this inverse problem–the measurements are taken in Fourier space but the constraints are in Object space.

3 Bayesian Estimation and Symmetry Breaking

A Bayesian estimator minimizes the conditional expected value of a cost function that is a function of both the estimated and the true values of the random variables. The conditioning is on the observations y_k.

Let ϕ_n be the random field and $\hat{\phi}_n$ be the estimate. I consider the l_2 cost function $\sum_n(\phi_n - \hat{\phi}_n)^2$ where the sum is over all lattice sites. (For binary lattice variables this is the same as the equal-penalty cost function $\sum_n(1 - \delta_{\phi_n,\hat{\phi}_n})$ where $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$). This cost function is natural and the solution of its minimization can be computed analytically. Specifically, the solution [14, 15] is to compute $\Pr(\phi_n = 1|\{y\}) = E(\phi_n|\{y\})$ and then set $\hat{\phi}_n$ to zero (respectively, one) if this probability is less than (respectively, greater than) one half. Therefore, it is necessary to compute the a posteriori expectation of the MRF.

In theory the needed expectations can be computed by summing $e^{-H'}$ or $\phi_n e^{-H'}$ over all configurations of the MRF lattice variables $\{\phi\}$. However, the Hamiltonian $H' = H^{\text{apriori}} + H^{\text{obs}}$ has too many symmetries to be useful in this Bayesian estimation problem. Specifically, in one dimension, if $\bar{\phi}_n$ is one configuration, described as a function of n, then $\bar{\phi}_{n+n_0}$ will

have exactly the same total energy H'. By summing over all of these shifted configurations, each with the same weight $e^{-H'}$, the resulting expectation will be constant, that is, it will have a constant value that will not depend on n. Similarly for reflections through the origin.

In order to solve this problem in general, it is necessary to break the unwanted symmetries of H'. In the one-dimensional case, for example, it is necessary to favor a particular translation n_0 over all other translations. A natural method to achieve this in general is to add an additional symmetry-breaking term $H^{\text{s.b.}}$ to the Hamiltonian where

$$H^{\text{s.b.}} = qL \sum_{n=0}^{L-1} \psi_{-n}\phi_n \tag{1}$$

and where ψ_n is real and periodic with period L. If $\psi_n = \begin{cases} 0 & \text{if } n = 0 \\ (L-n)/L & \text{if } n \neq 0 \end{cases}$ then $H^{\text{s.b.}}$ is the first moment of the field ϕ_n and the translational symmetry breaking effect is obvious[4]. This convolutional form for $H^{\text{s.b.}}$ is a "good" choice because it can be viewed as a mild perturbation (it depends only linearly on the field ϕ_n) and because, like the quadratic H^{apriori}, it can be written as a linear combination of functions of individual Fourier coefficients. While ψ_n could be set independently of the data, it is not clear what criteria should be employed. I take a different approach and use ψ_n as the parameters in a data-dependent adaptive estimation scheme.

The definition of the total Hamiltonian as $H = H^{\text{apriori}} + H^{\text{obs}} + H^{\text{s.b.}}$ and the choice of estimation goal form a complete definition of the problem. The remainder of this paper is devoted to the mathematics of calculating (approximations) to

$$Z^{\text{exact}}(\{y\}) = \sum_{\{\phi\}} e^{-H(\{y\},\{\phi\})} \tag{2}$$

$$E^{\text{exact}}(\phi_n|\{y\}) = \frac{1}{Z^{\text{exact}}(\{y\})} \sum_{\{\phi\}} \phi_n e^{-H(\{y\},\{\phi\})}. \tag{3}$$

The close relationship with calculations in statistical mechanics should be obvious.

[4]In higher dimensions the first moment is a vector. However, $H^{\text{s.b.}}$ must be a scalar.

4 Outline of the Calculation

The calculation proceeds in the following fashion. First, the entire calculation is done in terms of the coefficients of the Fourier series of the MRF field ϕ_n. This is the natural choice of variables because H^{obs}, which is quartic in the ϕ_n, is "diagonal" in this choice of variables. This is the reason for the care in choosing H^{apriori} and $H^{\text{s.b.}}$ as described above. Second, two approximations are introduced to address two different problems. First, the zero-one nature of the MRF lattice variables is very difficult to deal with. Therefore, the spherical model, which is a relaxation of this constraint, is introduced. Second, even with the spherical model, the problem has high dimensional exponential-of-quartic integrals which cannot be computed exactly. Therefore an asymptotic small noise approximation is introduced where the observation noise is assumed to have small variance. That is, in H^{obs} it is assumed that $\sigma_k \downarrow 0$. With these two approximations it is possible to compute the desired expectations analytically.

Two different asymptotic approximations are considered. In the first approximation ("Problem 1"), $\sigma_k \downarrow 0$ so that $H^{\text{obs}} \uparrow \infty$. Therefore, the a priori model H^{apriori} is progressively forgotten. In the second approximation ("Problem 2"), $H^{\text{apriori}} \uparrow \infty$ also, but $\frac{H^{\text{apriori}}}{H^{\text{obs}}} \rightarrow \chi$, a nonzero finite constant. In this case the a priori model never becomes insignificant.

In more detail, once the symmetry breaking term $H^{\text{s.b.}}$ has been introduced and $H = H^{\text{apriori}} + H^{\text{obs}} + H^{\text{s.b.}}$ has been defined, the calculation using the spherical model and asymptotic approximation precedes in the following fashion. The sums over the lattice variables are written as integrals over a singular measure and then the desired measure is approximated by a second, also singular, measure (Step 1). The spherical model is this change of measure. Specifically, instead of concentrating the measure at the corners of a hypercube representing the binary constraints on the lattice variables, the new measure weighs equally all points on a sphere circumscribed around the hypercube. The integrals are written in terms of Fourier coordinates (Step 2), the Fourier coordinates are written in terms of magnitude and rotated

phase variables (Step 3) which decouple in the spherical model, and the phase integrals are performed exactly in terms of modified Bessel functions (Step 4). The remaining integrals over the magnitudes are performed by the asymptotic approximation.

Two different asymptotic approximations are defined (Step 5). In both cases the integral is of Laplace type and the integration region is one orthant (the variables are all positive) of a manifold (from the spherical model constraint). The two different asymptotic approximations turn out to differ only in the definition of certain constants. Some notation is defined (Step 6) and some properties of the nonexponential portion of the integrand are noted (Step 7). The critical point can be computed explicitly (Step 8).

I give formal calculations rather than rigorous proofs of the asymptotic formulae. First the plan is outlined (Step 9). The plan depends on the implicit function theorem in order to deal with the manifold constraint, Taylor series expansions (around the critical point) of the exponent and of the nonexponential portion of the integrand, and expectations of polynomials of Gaussian random variables on the half (rather than full) line in order to deal with the orthant constraint. The Taylor expansion results are stated (Step 10). These results are complicated by the fact that the nonexponential part of the integrand typically vanishes at the critical point. Finally, the chain of approximations implied by the plan is applied in order to compute the formulae for the leading term of the asymptotic expansions (Step 11). I describe why rigorous proofs are difficult.

The actual conditional expectations are ratios of the asymptotic expansions (see, e.g., eqn. 6) where the critical point is the same in the numerator and denominator. This leads to major simplifications which are described (Step 12). Finally, the nonlinear thresholding to reconstruct the signal ϕ_n is described (Step 13).

5 Spherical Model

The summations of eqns. 2 and 3 over configurations of the binary-valued ϕ_n $n \in \{0, \ldots, L\}$ are written as integrals over R^L with a weighting function

$$w^{\text{exact}} = \prod_{n=0}^{L-1} \delta(\phi_n(\phi_n - 1))$$

where $\delta(x)$ is the Dirac delta function and $\delta(f(x))$ means

$$\delta(f(x)) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{jkf(x)} dk$$

in the distributional sense.

Let $\vec{\phi} = (\phi_1, \ldots, \phi_L)$. The spherical model approximation is to replace w^{exact}, which constrains $\vec{\phi}$ to lie at the corners of an L-dimensional hypercube, by $w^{\text{spherical}}$, which is defined to constrain $\vec{\phi}$ to lie on the hypersphere circumscribed around the hypercube.

Derivation of the hypersphere equation is simplified by considering $\bar{\phi}_n \in \{-1, +1\}$. In this case the center of the hypersphere is at the origin and its radius is $R = \sqrt{\sum_{n=0}^{L-1} 1^2} = \sqrt{L}$, and therefore its equation is $\sum_{n=0}^{L-1} \bar{\phi}_n^2 = L$. Since $\phi_n = \frac{1}{2}(\bar{\phi}_n + 1)$, the equation for the hypersphere for ϕ_n is $\sum_{n=0}^{L-1} \phi_n(\phi_n - 1) = 0$. Therefore, the spherical model weighting function is

$$w^{\text{spherical}} = \delta(\sum_{n=0}^{L-1} \phi_n(\phi_n - 1)).$$

It is difficult to assess the errors caused by the spherical model because few MRF problems can be exactly solved. The Ising model has site variables $\bar{\phi}_n \in \{\pm 1\}$ and, in zero external field, has Hamiltonian $H = -J \sum_{<i,j>} \bar{\phi}_i \bar{\phi}_j$ where the sum is over nearest neighbors. In two dimensions the exact solution is known and there is a critical point at temperature $\frac{2J}{kT_c} = .881$ while in the spherical model corresponding to the two dimensional Ising model there is no critical point (i.e., $\frac{2J}{kT_c} = \infty$). However, in three dimensions the approximation is much more accurate. Specifically, though the exact solution of the three dimensional Ising model is not known, the critical point is believed to lie near $\frac{2J}{kT_c} = .443$ while the spherical model corresponding to the three dimensional Ising model has a critical point at

$\frac{2J}{kT_c} = .505$. Note that these comparisons (from [16]) are for collective properties of infinite homogeneous lattices while this paper describes work concerning individual site statistics in finite inhomogeneous lattices. I am not aware of comparisons, presumably based in part on numeric simulation, which are more directly relevant to this paper. This completes Step 1.

6 Fourier Coordinates

Because H^{obs} is "diagonal" in the Fourier coefficients Φ_k of the field ϕ_n, the coefficients Φ_k are the natural variables for this problem. In this section, H and $w^{\text{spherical}}$ are rewritten in terms of these coordinates (Step 2) and magnitude and rotated phase variables are introduced (Step 3).

Define the double Fourier series expansion of w_2 as

$$W_2(k_1, k_2) = \sum_{n_1=0}^{L-1} \sum_{n_2=0}^{L-1} e^{-jn_1k_1\frac{2\pi}{L}} w_2(n_1, n_2) e^{-jn_2k_2\frac{2\pi}{L}}.$$

Properties of W_2 that follow from $w_2 \in R$ will be useful in what follows.

Using the definition of W_2, H^{apriori} can be written as

$$H^{\text{apriori}} = \frac{1}{L} \sum_{k=0}^{L-1} |\Phi_k|^2 W_2(-k, k) + w_1\Phi_0.$$

H^{obs} is already expressed in terms of Φ. Since $H^{\text{s.b.}}$ is a convolution evaluated at sample 0, $H^{\text{s.b.}}$ can be written

$$H^{\text{s.b.}} = q \sum_{k=0}^{L-1} \Psi_k\Phi_k$$

where Ψ_k are the Fourier coefficients of ψ_n. Finally, by using Parseval's Theorem and $\phi_n \in R$, $w^{\text{spherical}}$ can be written

$$w^{\text{spherical}} = \delta(\frac{1}{L} \sum_{k=0}^{L-1} |\Phi_k|^2 - \Phi_0).$$

The variables $\{\phi\}$ are real and therefore $\Phi_k = \Phi^*_{L-k}$. Assume that L, the number of lattice sites per unit cell, is odd. Then for any d (the dimension of the space) only Φ_0 is guaranteed to be real[5]. Furthermore, again because ϕ_n is real, not all of the Fourier coefficients Φ_k can

[5] For L even an increasing number of Φ_k are guaranteed to be real as d increases.

be taken as independent degrees of freedom. For L odd it is convenient to take

$$\begin{array}{ccc} \Re\Phi_0 & & \\ \Re\Phi_1 & , & \Im\Phi_1 \\ \vdots & & \vdots \\ \Re\Phi_{\frac{L-1}{2}} & , & \Im\Phi_{\frac{L-1}{2}} \end{array}$$

as independent. Define $\Phi_{r,k} = \Re\Phi_k$, $\Phi_{i,k} = \Im\Phi_k$, $K_L = \{0,1,\ldots,\frac{L-1}{2}\}$, and $K_L^+ = \{1,\ldots,\frac{L-1}{2}\}$.

Writing out the total Hamiltonian for the L odd case gives

$$\begin{aligned} H \;=\; & \frac{1}{2\sigma_0^2}y_0^2 \\ & + \Phi_{r,0}\left[w_1 + q\Psi_0\right] \\ & + \Phi_{r,0}^2\left[\frac{1}{L}W_2(0,0) + \frac{-1}{\sigma_0^2}y_0\right] \\ & + \Phi_{r,0}^4\frac{1}{2\sigma_0^2} \\ & + \sum_{k=1}^{\frac{L-1}{2}}\left\{\frac{1}{2\sigma_k^2}y_k^2 + \frac{1}{2\sigma_{L-k}^2}y_{L-k}^2\right. \\ & + \Re\{\Phi_k\Psi_k\}2q \\ & + |\Phi_k|^2\left[\frac{2}{L}W_2(-k,k) - \frac{1}{\sigma_k^2}y_k - \frac{1}{\sigma_{L-k}^2}y_{L-k}\right] \\ & \left. + |\Phi_k|^4\left[\frac{1}{2\sigma_k^2} + \frac{1}{2\sigma_{L-k}^2}\right]\right\} \end{aligned}$$

Similarly

$$w^{\text{spherical}} = \delta(\frac{1}{L}\Phi_0^2 - \Phi_0 + \frac{2}{L}\sum_{k=1}^{\frac{L-1}{2}}|\Phi_k|^2).$$

This completes Step 2.

Introduce a parameter β, analogous to inverse temperature in statistical mechanics, which allows the entire Hamiltonian to be simultaneously scaled. This scaling is analogous to scaling the inverse of the variance of a Gaussian distribution. The choice $\beta = 1$ leaves the variance as set by w_1, w_2 in H^{apriori}, σ_k^2 in H^{obs}, and q, w_n in $H^{\text{s.b.}}$ unchanged.

In H the contribution of different Φ_k is additive. Define

$$-\beta h_k = \begin{cases} a_{0,0} + a_{0,1}\Phi_{r,0} + a_{0,2}\Phi_{r,0}^2 + a_{0,4}\Phi_{r,0}^4 & \text{if } k = 0 \\ a_{k,0} + a_{k,1}\Re\{\Phi_k \frac{\Psi_k}{|\Psi_k|}\} + a_{k,2}|\Phi_k|^2 + a_{k,4}|\Phi_k|^4 & \text{if } k \neq 0 \end{cases}$$

where the $a_{k,j}$ have the straightforward definitions stated in Appendix A. Then,

$$-\beta H = \sum_{k=0}^{\frac{L-1}{2}} -\beta h_k.$$

For $k \in K_L^+$, change variables to $\Phi_k' = \frac{\Psi_k}{|\Psi_k|}\Phi_k$. Introduce magnitude and phase variables by

$$r_k = \begin{cases} \Phi_{r,0} & \text{if } k = 0 \\ |\Phi_k'| & \text{if } k \in K_L^+ \end{cases}$$

$$\theta_k = \begin{cases} 0 & \text{if } k = 0 \\ \angle\Phi_k' & \text{if } k \in K_L^+ \end{cases}$$

where the fact that Φ_0 is real has been used.

In these variables $-\beta h_k$ has the form

$$-\beta h_k = \begin{cases} a_{0,0} + a_{0,1}r_0 + a_{0,2}r_0^2 + a_{0,4}r_0^4 & \text{if } k = 0 \\ a_{k,0} + a_{k,1}r_k\cos\theta_k + a_{k,2}r_k^2 + a_{k,4}r_k^4 & \text{if } k \in K_L^+ \end{cases}$$

$$w^{\text{spherical}} = \delta(\frac{1}{L}r_0^2 - r_0 + \frac{2}{L}\sum_{k=1}^{\frac{L-1}{2}} r_k^2).$$

Note that $w^{\text{spherical}}$ depends only on the r_k variables and is independent of the θ_k variables.

In preparation for the angular integrations, the Hamiltonian is divided into two parts, one part ($h_{k,\theta,r}$) that includes all of the θ dependence and some r dependence also, and a second part ($h_{k,r}$) that contains no θ dependence. The definitions are

$$-\beta h_{k,\theta,r} = \begin{cases} 0 & \text{if } k = 0 \\ a_{k,1}r_k\cos\theta_k & \text{if } k \in K_L^+ \end{cases}$$

$$-\beta h_{k,r} = \begin{cases} a_{0,0} + a_{0,1}r_0 + a_{0,2}r_0^2 + a_{0,4}r_0^4 & \text{if } k = 0 \\ a_{k,0} + a_{k,2}r_k^2 + a_{k,4}r_k^4 & \text{if } k \in K_L^+ \end{cases}$$

Then, $-\beta h_k = -\beta h_{k,\theta,r} - \beta h_{k,r}$. This completes Step 3.

7 Partition Function and Moments

Combining the results presented in the previous two sections, in this section the approximations under the spherical model to the partition function $Z^{\text{exact}}(\{y\})$ (eqn. 2) and moments $E^{\text{exact}}(\phi_n|\{y\})$ (eqn. 3) are written out and the integrals over the rotated phase coordinates (Step 4) are performed. The partition function Z after the introduction of the spherical model is

$$\begin{aligned} Z(\{y\}) &= \int_{-\infty}^{+\infty} d\Phi_{r,0} \int_{-\infty}^{+\infty} d\Phi_{r,1} \int_{-\infty}^{+\infty} d\Phi_{i,1} \cdots \int_{-\infty}^{+\infty} d\Phi_{r,\frac{L-1}{2}} \int_{-\infty}^{+\infty} d\Phi_{i,\frac{L-1}{2}} w^{\text{spherical}} e^{-\beta H} \\ &= \int_{-\infty}^{+\infty} dr_0 \int_0^{+\infty} r_1 dr_1 \int_0^{2\pi} d\theta_1 \cdots \int_0^{+\infty} r_{\frac{L-1}{2}} dr_{\frac{L-1}{2}} \int_0^{2\pi} d\theta_{\frac{L-1}{2}} w^{\text{spherical}} e^{-\beta H} \\ &= \int_{-\infty}^{+\infty} dr_0 \int_0^{+\infty} dr_1 \cdots \int_0^{+\infty} dr_{\frac{L-1}{2}} w^{\text{spherical}} \exp\Big(\sum_{k=0}^{\frac{L-1}{2}} -\beta h_{k,r}\Big) \prod_{k=1}^{\frac{L-1}{2}} r_k \Theta_k^0(r_k) \end{aligned} \tag{4}$$

where for $k \in K_L^+$

$$\begin{aligned} \Theta_k^0(r_k) &= \int_0^{2\pi} d\theta_k \exp(-\beta h_{k,\theta,r}) \\ &= 2\pi I_0(a_{k,1} r_k). \end{aligned} \tag{5}$$

The θ_k integral is standard [17, eqn. (8.431 4.) for $\nu = 0$].

In accord with the use of Fourier variables, the mean of the field is computed in terms of the mean of its Fourier coefficients. That is, an approximation is computed under the spherical model to $E^{\text{exact}}(\Phi_k|\{y\})$ rather than to $E^{\text{exact}}(\phi_n|\{y\})$. For the mean of Φ_k, the integrand for Z is multiplied by

$$\Phi_k = \begin{cases} r_0 & \text{if } k = 0 \\ r_k \exp \theta_k \frac{\Psi_k^*}{|\Psi_k|} & \text{if } k \in K_L^+ \end{cases}$$

and the result is scaled by $\frac{1}{Z}$. Therefore,

$$\begin{aligned} E(\Phi_{r,0}|\{y\}) &= \frac{1}{Z(\{y\})} \int_{-\infty}^{+\infty} dr_0 \int_0^{+\infty} dr_1 \cdots \int_0^{+\infty} dr_{\frac{L-1}{2}} \times \\ &\quad \times r_0 w^{\text{spherical}} \exp\Big(\sum_{k=0}^{\frac{L-1}{2}} -\beta h_{k,r}\Big) \prod_{k=1}^{\frac{L-1}{2}} r_k \Theta_k^0(r_k) \end{aligned} \tag{6}$$

$$E(\Phi_{i,0}|\{y\}) = 0 \tag{7}$$

$$E(\Phi_k|\{y\}) = \frac{1}{Z(\{y\})}\int_{-\infty}^{+\infty} dr_0 \int_0^{+\infty} dr_1 \cdots \int_0^{+\infty} dr_{\frac{L-1}{2}} \times$$
$$\times \frac{\Psi_k^*}{|\Psi_k|} r_k w^{\text{spherical}} \exp\Big(\sum_{l=0}^{\frac{L-1}{2}} -\beta h_{l,r}\Big) r_k \Theta_k^1(r_k) \prod_{\substack{l=1 \\ l\neq k}}^{\frac{L-1}{2}} r_l \Theta_l^0(r_l)$$
$$k \in K_L^+ \tag{8}$$

where for $k \in K_L^+$

$$\Theta_k^1(r_k) = \int_0^{2\pi} d\theta_k \exp(j\theta_k) \exp(-\beta h_{k,\theta,r})$$
$$= 2\pi I_1(a_{k,1} r_k). \tag{9}$$

The θ_k integral is the derivative with respect to the parameter (justified by the Lebesgue dominated convergence theorem) of a standard integral [17, eqn. (8.431 1.) for $\nu = 0$].

The remaining $E(\Phi_k|\{y\})$ are specified by $\Phi_k = \Phi_{L-k}^*$, that is, $E(\Phi_k|\{y\}) = E(\Phi_{L-k}|\{y\})^*$. This completes Step 4.

8 Asymptotics

Unfortunately, the magnitude integrals presented in the previous section (i.e., eqns. 4, 6, 7, and 8) do not appear to be solvable in terms of standard functions. In the Bayesian context, especially considering the relatively good accuracy of the crystallographic data, it is natural to consider an asymptotic evaluation in terms of small variances of the observation noise. The parameters that one typically considers for asymptotics in statistical mechanics problems are less appropriate. For instance, asymptotics in the lattice spacing/number of lattice sites would increase without bound the number of random variables being estimated while asymptotics in the "external field" strengths corresponds to asymptotics in the scattering intensities, which need not be small.

Two different asymptotic limits are considered. One limit, denoted Problem 1, is purely a small observation noise limit. That is, these integrals are evaluated in the limit $\sigma_k^2 \downarrow 0$. More

precisely, the assumption is that $\sigma_k^2 = \frac{1}{\lambda}\bar{\sigma}_k^2$ and $\lambda \uparrow \infty$. The second limit, denoted Problem 2, combines the small observation noise limit with a proportional scaling of the a priori Hamiltonian. Specifically, the assumption is that $\sigma_k^2 = \frac{1}{\lambda}\bar{\sigma}_k^2$, $W_2(k_1, k_2) = \lambda\chi\bar{W}_2(k_1, k_2)$, $w_1 = \lambda\chi\bar{w}_1$, $\lambda \uparrow \infty$, and χ is a fixed real number.

Two different problems are formulated because in Problem 1, the true small observation noise limit, the influence of the a priori portion of the Hamiltonian relative to the observational portion of the Hamiltonian decreases as λ grows. Though the resulting estimators are used at finite λ, they are derived in the $\lambda \to \infty$ limit and therefore may undesirably suppress the prior knowledge represented by the a priori portion of the Hamiltonian. On the other hand in Problem 2 the a priori portion is rescaled so that the a priori and observational portions of the Hamiltonian have constant (in the sense of fixed ratio) influence. This completes Step 5.

In Problem 1 in the case when no observation is taken at frequency k, there is no λ dependence in $-\beta h_k$. However, $w^{\text{spherical}}$ continues to couple this r_k integral to the other r_l integrals, some of which must have λ dependence.

Both Problems 1 and 2 concern the asymptotic expansion of integrals of the form $\int_D \alpha(x)e^{\lambda\gamma(x)}dx$ in the limit $\lambda \to \infty$ where γ is real and therefore the integral is of Laplace type [18]. Not only the order in λ but also the numerical coefficient of the first nonzero term in the $\lambda \to \infty$ asymptotic series is required.

The points where the exponent γ attains a global maximum, called critical points, play an important role in the large-λ asymptotics because as $\lambda \to \infty$ the entire contribution to the integral comes from a neighborhood of those points. Though it does not contribute to the determination of the critical points, the behavior of α (the nonexponential part of the integrand), especially the points at which α and perhaps its derivatives vanish, is also important because these points may, and in fact do, occur at the critical points. As will be described, the problem is difficult because the critical point lies on the boundary of the domain of integration D, the boundary of D is not smooth at the critical point, and α

vanishes to high and data-dependent order at the critical point.

9 Asymptotics–Notation

In this section the first goal is to define notation so that the partition function (eqn. 4) and conditional means (eqns. 6, 7, and 8) can be written

$$\begin{aligned} Z(\lambda) &= \int g_Z w^{\text{spherical}} e^{-\lambda\beta H_\lambda} \\ E(\Phi_{r,0}|\{y\})(\lambda) &= \int g_0 w^{\text{spherical}} e^{-\lambda\beta H_\lambda} \\ E(\Phi_k|\{y\})(\lambda) &= \int g_k w^{\text{spherical}} e^{-\lambda\beta H_\lambda} \quad k \in K_L^+. \end{aligned}$$

First define some quantities related to the exponent. Having introduced $h_{k,\theta,r}$ and $h_{k,r}$ in Section 7 and λ and χ in Section 8, it is helpful to have a second set of constants that show the dependencies more explicitly than the $a_{k,n}$. Define $b_{k,ns}$ where n is the order of the Φ dependence and s is a suffix. The three suffixes are $s = a$ for dependence on σ_k (which automatically implies dependence on λ), $s = b$ for dependence on λ but not σ (this can only occur in Problem 2), and $s = c$ for no dependence on λ. Because $h_{k,\theta,r}$ and $h_{k,r}$ have different order of dependence on Φ_k, a given $b_{k,ns}$ constant automatically enters into one or the other but not both.

The two sets of $b_{k,ns}$ definitions, one for Problem 1 and one for Problem 2, are in Appendix A. The only difference between Problem 1 and 2 is the definition of these constants $b_{k,ns}$ and for both Problem 1 and Problem 2 the $a_{k,n}$ in terms of the $b_{k,ns}$ take the form

$$\begin{aligned} a_{k,0} &= -\lambda b_{k,0a} \\ a_{k,1} &= \lambda b_{k,1b} + b_{k,1c} \\ a_{k,2} &= \lambda b_{k,2a} + \lambda b_{k,2b} + b_{k,2c} \\ a_{k,4} &= -\lambda b_{k,4a}. \end{aligned}$$

Make explicit the λ dependence of the exponent by defining

$$-\beta h_{k,r0} = \begin{cases} b_{0,1c}r_0 + b_{0,2c}r_0^2 & \text{if } k = 0 \\ b_{k,2c}r_k^2 & \text{if } k \in K_L^+ \end{cases} \tag{10}$$

$$-\beta h_{k,r1} = \begin{cases} -b_{0,0a} + b_{0,1b}r_0 + (b_{0,2a} + b_{0,2b})r_0^2 - b_{0,4a}r_0^4 & \text{if } k = 0 \\ -b_{k,0a} + (b_{k,2a} + b_{k,2b})r_k^2 - b_{k,4a}r_k^4 & \text{if } k \in K_L^+ \end{cases}$$

so that

$$-\beta h_{k,r} = -\beta h_{k,r0} - \lambda \beta h_{k,r1}$$

and there is no other λ dependence in $h_{k,r}$. Define

$$-\beta H_\lambda = \sum_{k=0}^{\frac{L-1}{2}} -\beta h_{k,r1}.$$

Second, define some quantities related to the nonexponential part of the integrand. Define

$$q_k^0(x) = \begin{cases} \exp(-\beta h_{0,r0}(x)) & k = 0 \\ x\Theta_k^0(x)\exp(-\beta h_{k,r0}(x)) & k \in K_L^+ \end{cases} \tag{11}$$

$$q_k^1(x) = \begin{cases} x\exp(-\beta h_{0,r0}(x)) & k = 0 \\ \frac{\Psi_k^*}{|\Psi_k|}x^2\Theta_k^1(x)\exp(-\beta h_{k,r0}(x)) & k \in K_L^+ \end{cases} \tag{12}$$

$$q_k^{0(n)}(x) = \frac{d^n q_k^0(x)}{dx^n}$$

$$q_k^{1(n)}(x) = \frac{d^n q_k^1(x)}{dx^n}.$$

Then

$$gz(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = \prod_{k=0}^{\frac{L-1}{2}} q_k^0(r_k) \tag{13}$$

$$g_k(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = q_k^1(r_k) \prod_{\substack{j=0 \\ j\neq k}}^{\frac{L-1}{2}} q_j^0(r_j) \quad k \in K_L \tag{14}$$

which are all independent of λ.

The second goal is to fix some notation concerning the critical point. Let $\rho \in R^{\frac{L-1}{2}+1}$. $\rho = (\rho_0, \rho_1, \ldots, \rho_{\frac{L-1}{2}})$ be the critical point, and define $\bar{\rho} \in R^{\frac{L-1}{2}}$, $\bar{\rho} = (\rho_1, \ldots, \rho_{\frac{L-1}{2}})$. Similarly, the variable r will always denote a variable in $R^{\frac{L-1}{2}+1}$ while the variable $\bar{r}$ will always denote

a variable in $R^{\frac{L-1}{2}}$. Components of the critical point ρ that are zero play an important role. Define

$$A_\rho = \{k \in K_L | \rho_k = 0\} \tag{15}$$

$$\bar{A}_\rho = \{k \in K_L^+ | \rho_k = 0\}. \tag{16}$$

Define

$$C(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = \frac{1}{L} r_0^2 - r_0 + \frac{2}{L} \sum_{k=1}^{\frac{L-1}{2}} r_k^2$$

so that

$$w^{\text{spherical}} = \delta(C(r_0, r_1, \ldots, r_{\frac{L-1}{2}})).$$

Therefore, the integrals of eqns. 4, 6, 7, and 8 are only over (a subset of) the manifold defined by $C(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = 0$. The implicit function theorem assures the existence in a neighborhood of ρ of a continuously differentiable function $\eta_\rho : R^{\frac{L-1}{2}} \rightarrow R$ such that $C(\eta_\rho(\bar{r}), \bar{r}) = 0$ in this neighborhood assuming that $(\partial_{r_0} C)(\rho) = \frac{2}{L}\rho_0 - 1 \neq 0$ which is true so long as $\rho_0 \neq \frac{L}{2}$. For notational convenience define

$$F_\rho : R^{\frac{L-1}{2}} \rightarrow R^{\frac{L-1}{2}+1}$$

$$(F_\rho(\bar{r}))_k = \begin{cases} \eta_\rho(\bar{r}) & \text{if } k = 0 \\ r_k & \text{if } k \in K_L^+ \end{cases}. \tag{17}$$

Note that $F_\rho(\bar{\rho}) = \rho$. This completes Step 6.

The third goal is to state properties of the zeros of g and the derivatives of g. These properties are stated in terms of the corresponding properties of the q functions. For $k = 0$ (respectively $k \in K_L^+$) the behavior of the q functions in the region $x \in R$ (respectively $x \geq 0$) is important. In this region, elementary computations reveal that

$$q_0^0(x) > 0 \quad \forall x \in R$$

$$q_k^0(x) > 0 \quad \forall x > 0, \quad q_k^0(0) = 0 \quad k \in K_L^+$$

$$q_0^1(x) > 0 \quad \forall x > 0, \quad q_0^1(x) < 0 \quad \forall x < 0, \quad q_0^1(0) = 0$$

$$
\begin{aligned}
q_k^1(x) &> 0 \quad \forall x > 0, \quad q_k^1(x) = 0 \quad k \in K_L^+ \\
q_k^{0(1)}(0) &\neq 0 \quad k \in K_L^+ \\
q_0^{1(1)}(0) &\neq 0 \\
q_k^{1(1)}(0) &= 0 \quad k \in K_L^+ \\
q_k^{1(2)}(0) &= 0 \quad k \in K_L^+ \\
q_k^{1(3)}(0) &\neq 0 \quad k \in K_L^+.
\end{aligned} \tag{18}
$$

This completes Step 7.

Finally, Gaussian integrals play an important role. Define

$$
\begin{aligned}
N &: R^{n \times n} \rightarrow R \\
N(Q) &= \sqrt{\det(\frac{1}{2\pi} Q)}
\end{aligned}
$$

which is the normalization factor for a Gaussian density with covariance matrix Q^{-1} (i.e., $p_{m,Q^{-1}}(r) = N(Q)\exp(-\frac{1}{2}(r-m)^T Q(r-m))$).

10 Asymptotics–Critical Point

The critical point is computed using standard techniques from constrained optimization theory [19, 20].

Consider two optimization problems:

$$
\begin{aligned}
\text{Opt 1} &: \min \beta H_\lambda \\
&\quad \text{subject to } C = 0, r_k \geq 0 \; k \in K_L^+ \\
\text{Opt 2} &: \min \beta H_\lambda \\
&\quad \text{subject to } C = 0.
\end{aligned}
$$

The development described in previous sections leads to problems of the type Opt 1. The solution of such problems appears to require the use of Kuhn-Tucker theory because of the

presence of the inequality constraints on r_k $k \in K_L^+$. However, because βH_Λ depends on r_k $k \in K_L^+$ only through r_k^2, any extreme point in the complement of $\{r_0 \in R\} \times \{r_k \geq 0 | k \in K_L^+\}$ maps to an extreme point within $\{r_0 \in R\} \times \{r_k \geq 0 | k \in K_L^+\}$ which has the same value. Furthermore, there are no extreme points on the boundary for Opt 1 that are not also extreme points for Opt 2. Therefore, assuming that the solution is suitably reflected into the orthant $\{r_0 \in R\} \times \{r_k \geq 0 | k \in K_L^+\}$, it is sufficient to solve Opt 2, which requires only the simpler Lagrange theory.

The Lagrange multiplier variable is denoted τ. All points are regular points. The application of the Lagrange conditions yields the following sets of equations. The $k = 0$ component of the gradient condition gives

$$-b_{0,1b} - 2(b_{0,2a} + b_{0,2b})\rho_0 + 4b_{0,4a}\rho_0^3 + \tau(\frac{2}{L}\rho_0 - 1) = 0 \tag{19}$$

while the $k \in K_L^+$ components give

$$-2(b_{k,2a} + b_{k,2b})\rho_k + 4b_{k,4a}\rho_k^3 + \tau\frac{4}{L}\rho_k = 0. \tag{20}$$

The constraint condition gives

$$\frac{1}{L}\rho_0^2 - \rho_0 + \frac{2}{L}\sum_{k=1}^{\frac{L-1}{2}} \rho_k^2 = 0. \tag{21}$$

The subspace

$$M(\rho) = \{y | \nabla C(\rho)^T y = 0\}$$

simplifies to

$$M(\rho) = \{y | (\frac{2}{L}\rho_0 - 1)y_0 + \frac{4}{L}\sum_{k=1}^{\frac{L-1}{2}} \rho_k y_k = 0\}.$$

Therefore, the second order condition gives

$$y^T \operatorname{diag}(-2(b_{0,2a} + b_{0,2b}) + 12b_{0,4a}\rho_0^2 + \tau\frac{2}{L},$$
$$\ldots, -2(b_{k,2a} + b_{k,2b}) + 12b_{k,4a}\rho_k^2 + \tau\frac{4}{L}, \ldots)y \geq 0 \quad y \in M(\rho)$$

or equivalently

$$0 \leq \left[-2(b_{0,2a}+b_{0,2b})+12b_{0,4a}\rho_0^2+\tau\frac{2}{L}\right]y_0^2 + \sum_{k=1}^{\frac{L-1}{2}}\left[-2(b_{k,2a}+b_{k,2b})+12b_{k,4a}\rho_k^2+\tau\frac{4}{L}\right]y_k^2 \quad y \in M(\rho). \tag{22}$$

The goal is to find all solutions ρ, τ of eqns. 19, 20 $k \in K_L^+$, and 21 satisfying eqn. 22. The approach is to solve the gradient equation (eqn. 20) for each $k \in K_L^+$ to get ρ_k as a function of τ. The function depends on whether an observation is or is not taken at frequency k. In addition, in order to get a single valued function, the second order condition (eqn. 22) must be taken into consideration. Then the pair of equations eqns. 19 and 21, after substitution of ρ_k as a function of τ $k \in K_L^+$, determine ρ_0 and τ. Finally, given τ, the expressions for ρ_k $k \in K_L^+$ as a function of τ fix the remaining ρ_k.

The value of ρ_k as a function of τ for a given frequency k depends on whether or not an observation was taken at frequency k. Define $B = \{k \in K_L^+ |$ an observation was taken at frequency $k\}$.

Consider the gradient equation (eqn. 20) at $k \neq 0$. First consider the case where an observation is taken at frequency k, i.e., $k \in B$. Then the three values

$$\rho_k(\tau) = 0, \pm\sqrt{\frac{(b_{k,2a}+b_{k,2b})-\tau\frac{2}{L}}{2b_{k,4a}}} \quad k \in K_L^+ \cap B$$

are the only alternatives for ρ_k. The negative square root need not be considered. Furthermore, the fact that $\rho_k \in R$ requires

$$(b_{k,2a}+b_{k,2b})-\tau\frac{2}{L} \geq 0$$

if the positive square root solution is to be acceptable. Therefore,

$$\rho_k(\tau) = \begin{cases} 0, \sqrt{\frac{(b_{k,2a}+b_{k,2b})-\tau\frac{2}{L}}{2b_{k,4a}}} & \text{if } \tau < \frac{L}{2}(b_{k,2a}+b_{k,2b}) \\ 0 & \text{if } \tau \geq \frac{L}{2}(b_{k,2a}+b_{k,2b}) \end{cases}$$

Suppose $\tilde{\rho}$, $\tilde{\tau}$ satisfied the gradient conditions ($k \in K_L$) and the constraint condition. Consider the second order condition. Suppose there is a $\tilde{k} \in B$ such that $\tilde{\rho}_{\tilde{k}} = 0$. Let

$\tilde{y} \in R^{\frac{L-1}{2}+1}$ be the vector with exactly one 1 in position $\tilde{k}$. That is, $(\tilde{y})_k = \delta_{k,\tilde{k}}$. Because $\tilde{\rho}_{\tilde{k}} = 0$, $\tilde{y} \in M(\tilde{\rho})$. Then the second order condition requires

$$0 \leq -2(b_{\tilde{k},2a} + b_{\tilde{k},2b}) + \tilde{\tau}\frac{4}{L}.$$

Therefore,

$$\tilde{\tau} \geq \frac{L}{2}(b_{\tilde{k},2a} + b_{\tilde{k},2b}).$$

Therefore, for ρ, τ that are local minima it is necessary to have

$$\rho_k(\tau) = \begin{cases} \sqrt{\frac{(b_{k,2a}+b_{k,2b})-\tau\frac{2}{L}}{2b_{k,4a}}} & \text{if } \tau < \frac{L}{2}(b_{k,2a} + b_{k,2b}) \\ 0 & \text{if } \tau \geq \frac{L}{2}(b_{k,2a} + b_{k,2b}) \end{cases} \quad k \in K_L^+ \cap B. \tag{23}$$

Second, continuing with the gradient equation (eqn. 20) at $k \neq 0$, consider the case where an observation is not taken at frequency k, i.e., $k \in K_L^+ - B$. In this case, $b_{k,0a} = b_{k,2a} = b_{k,4a} = 0$. Then the values

$$\rho_k(\tau) = \begin{cases} 0 & \text{if } \tau \neq \frac{L}{2}b_{k,2b} \\ \text{nonnegative} & \text{if } \tau = \frac{L}{2}b_{k,2b} \end{cases} \quad k \in K_L^+ - B$$

are the only alternatives for ρ_k.

Suppose $\tilde{\rho}$, $\tilde{\tau}$ satisfied the gradient conditions ($k \in K_L$) and the constraint condition. Consider the second order condition. Suppose there is a $\tilde{k} \in K_L^+ - B$ such that $\tilde{\rho}_{\tilde{k}} = 0$. Let $\tilde{y} \in R^{\frac{L-1}{2}+1}$ be the vector with exactly one 1 in position $\tilde{k}$. That is, $(\tilde{y})_k = \delta_{k,\tilde{k}}$. Because $\tilde{\rho}_{\tilde{k}} = 0$, $\tilde{y} \in M(\tilde{\rho})$. Then the second order condition requires

$$0 \leq -2b_{\tilde{k},2b} + \tilde{\tau}\frac{4}{L}.$$

Therefore,

$$\tilde{\tau} \geq \frac{L}{2}b_{\tilde{k},2b}.$$

Recall that for $\tilde{k} \in K_L^+ - B$ such that $\rho_{\tilde{k}} \neq 0$, there is already the requirement that $\tilde{\tau} = \frac{L}{2}b_{\tilde{k},2b}$. (There can be at most one such $\tilde{k}$). Therefore, for ρ, τ that are local minima it is

necessary to have

$$\begin{aligned} \tau &\geq \frac{L}{2} b_{k,2b} \; \forall k \in K_L^+ - B \\ \rho_k(\tau) &= \begin{cases} 0 & \text{if } \tau > \frac{L}{2} b_{k,2b} \\ \text{nonnegative} & \text{if } \tau = \frac{L}{2} b_{k,2b} \end{cases} \quad k \in K_L^+ - B. \end{aligned} \tag{24}$$

The equations eqns. 23 and 24 determine $\rho_k \; k \in K_L^+$ as a function of τ. Because of the step-like dependence of $\rho_k(\tau)$ on τ, it is most straightforward to solve the remaining two equations (eqns. 19 and 21) by partitioning the set of allowed τ values, which is R, at the discontinuities. The locations of the discontinuities depend on the measured data. Then, by hypothesizing that τ falls between some pair of adjacent discontinuities, one can derive a quadratic or cubic equation for ρ_0 from the constraint equation (eqn. 21). Using this value in the $k = 0$ term of the gradient equation (eqn. 19) allows the computation of τ, which may or may not fall into the hypothesized range of values. If τ does fall into the hypothesized range of values then it is straightforward to compute ρ_k and βH_λ for this local minimum. Finally, among all the local minimum of βH_λ, the critical point ρ is the point where βH_λ attains its global minimum (or equivalently the exponent $-\beta H_\lambda$ attains its global maximum). Because the partitioning of R involves only $\frac{L-1}{2}$ points, this is a very practical algorithm. See [13] for details. A very important point is that multiple components of the critical point will typically have value zero. This completes Step 8.

11 Asymptotics–Formulae

I give formal calculations in the spirit of [18] rather than rigorous proofs of the necessary formulae.

As discussed in Section 8, the goal is to compute the numerical value of the leading nonzero term of an asymptotic expansion of the integral

$$I(\lambda) = \int_{r_0=-\infty}^{+\infty} dr_0 \int_{r_1=0}^{\infty} dr_1 \ldots \int_{r_{\frac{L-1}{2}}=0}^{\infty} dr_{\frac{L-1}{2}} g(r) \delta(C(r)) e^{-\lambda \beta H_\lambda(r)}$$

as $\lambda \to \infty$ where g is any of g_Z, g_0, and g_k $k \in K_L^+$. That is, the goal is to compute an explicit formula $I_0(\lambda)$, i.e., a formula without integral signs and so forth, such that

$$\lim_{\lambda \to \infty} \frac{I(\lambda)}{I_0(\lambda)} = 1.$$

In the limit $\lambda \to \infty$ the contribution to $I(\lambda)$ comes from vanishing neighborhoods of the global maxima of $-\beta H_\lambda$ (the critical points) since outside of this region the integral is decreased by a factor $\exp(-\lambda\beta(H_\lambda(r) - H_\lambda(\rho)))$ and λ is arbitrarily large. Consider the contribution from one such critical point ρ and contract the region of integration to a neighborhood of ρ. If a component ρ_k $k \in K_L^+$ is zero, then ρ will lie on the boundary of the neighborhood. Throughout, by taking λ sufficiently large, it is possible to take the neighborhood as small as desired. Within the neighborhood of ρ, the inverse function theorem guarantees that the constraint $C(r) = 0$ can be solved for r_0 as a function of $r_1, \ldots, r_{\frac{L-1}{2}}$ (i.e., $\eta_\rho(\bar{r})$ and $F_\rho(\bar{r})$ in eqn. 17). Using this relationship the r_0 integration[6] can be performed leaving the integrand

$$g(F_\rho(\bar{r}))\exp(-\lambda\beta H_\lambda(F_\rho(\bar{r}))).$$

Also, within a neighborhood of the critical point ρ, the term $\beta H_\lambda(F_\rho(\bar{r}))$ can be replaced by a two term Taylor series expansion around $\bar{r} = \bar{\rho}$ and the term $g(F_\rho(\bar{r}))$ can be replaced by the least-order Taylor series expansion around $\bar{r} = \bar{\rho}$ that satisfies the following two conditions. First, it must have nonzero coefficients. Second, multiplication by the exponential of the two-term Taylor series expansion of $\beta H_\lambda(F_\rho(\bar{r}))$ and integration over an appropriate region must give a nonzero result. These two conditions result in computing the leading nonzero (rather than possibly zero) term. It is at this point that the vanishing of the nonexponential part of the integrand matters.

Finally, the region of integration can be expanded as follows since the additional contribution to the integral is negligible. For coordinates r_k $k \in K_L^+$ such that $\rho_k \neq 0$ the region

[6] Under suitable limitations on g one has $\int f(x)\delta(g(x))dx = \frac{f(g^{-1}(0))}{g'(g^{-1}(0))}$. The denominator is common to all integrals computed in this paper and will cancel from the ratios of interest. Therefore it is routinely dropped without comment.

of integration is expanded from the neighborhood of ρ_k to $(-\infty, +\infty)$ while, if $\rho_k = 0$, the region of integration is expanded from the neighborhood of ρ_k to $[0, +\infty)$. The two cases occur because if $\rho_k = 0$ then the integrand is not small at $r_k = 0^-$ and expansion of the region of integration for r_k into the negative half line gives large contributions that are not present in the original integral. Note that the resulting integral is in the form of the moment of a polynomial (from the Taylor series expansion of $g(F_\rho(\bar{r}))$) with respect to a Gaussian density (from the exponential of the Taylor series expansion of $-\beta H_\lambda(F_\rho(\bar{r}))$) where the moment is over a hyperplane in R^L rather than all of R^L due to the $\rho_k = 0$ coordinates. This completes Step 9.

The Taylor series around $\bar{\rho}$ of the exponent $-\beta H_\lambda(F_\rho(\bar{r}))$ has the form

$$-\beta H_\lambda(F_\rho(\bar{r})) \approx -\beta H_\lambda(\rho) - \frac{1}{2}(\bar{r} - \bar{\rho})^T L_\rho(\bar{r} - \bar{\rho}).$$

The absence of a linear term is due to the fact that ρ, which is an extreme point of the constrained optimization problem Opt 1, is also a stationary point. The form of L_ρ, an elementary calculation, is given later in this section.

The Taylor series around ρ of the nonexponential part of the integrand has several possible forms due to the fact that g can vanish at the critical point.

Case I: assume $g(\rho) \neq 0$. Then

$$g(F_\rho(\bar{r})) \approx g(\rho).$$

From the properties of g implied by eqn. 18, this requires that $\bar{A}_\rho = \emptyset$ (see eqn. 16).

Case II: assume $g(\rho) = 0$. There are subcases. First note that $\rho_0 = 0$ implies $\rho_k = 0$ $\forall k \in K_L^+$ by the constraint equation. The subcases are based on whether there is a unique derivative of lowest order of $g \circ F_\rho$ (function composition) that is nonzero at $\bar{\rho}$ versus several derivatives of the same order. The subcases are:

Case IIA (unique derivative):

assume $g(\rho) = 0$, $g \in \{g_Z, g_0, g_k\}$, $\rho_0 \neq 0$, $\rho_k = 0$ $\forall k \in \bar{A}_\rho \neq \emptyset$ or

assume $g(\rho) = 0$, $g \in \{g_Z, g_k\}$, $\rho_0 = 0$ implies $\rho_k = 0\ \forall k \in \bar{A}_\rho = K_L^+$;

Case IIB (nonunique derivative):

assume $g(\rho) = 0$, $g = g_0$, $\rho_0 = 0$ implies $\rho_k = 0\ \forall k \in \bar{A}_\rho = K_L^+$.

Since I believe that cases with $\rho_0 = 0$ are rare and have not seen one in simulation, and since the Case IIB subset of the $\rho_0 = 0$ cases is complicated, I will not present Case IIB here. For Case IIA the minimal Taylor series expansion is

$$g(F_\rho(\bar{r})) \approx \begin{cases} (\partial_{\bar{A}_\rho} g_Z)(\rho) \prod_{j \in \bar{A}_\rho} r_j & \text{if } g = g_Z \\ (\partial_{\bar{A}_\rho} g_0)(\rho) \prod_{j \in \bar{A}_\rho} r_j & \text{if } g = g_0 \\ (\partial_{\bar{A}_\rho} g_k)(\rho) \prod_{j \in \bar{A}_\rho} r_j & \text{if } g = g_k\ k \in K_L^+\ k \notin \bar{A}_\rho \\ \frac{1}{3!}(\partial_{r_k} \partial_{r_k} \partial_{\bar{A}_\rho} g_k)(\rho) r_k^3 \prod_{j \in \bar{A}_\rho - \{k\}} r_j & \text{if } g = g_k\ k \in K_L^+\ k \in \bar{A}_\rho \end{cases}$$

where $\partial_S = \prod_{i \in S} \partial_{r_i}$ and ∂_{r_i} is the partial derivative with respect to r_i. This completes Step 10.

These ideas lead to the follow chains of approximations.

$$\begin{aligned} I(\lambda) &= \int_{r_0 \in (-\infty, +\infty)} dr_0 \prod_{j=1}^{\frac{L-1}{2}} \int_{r_j \in [0,\infty)} dr_j g(r) \delta(C(r)) e^{-\lambda \beta H_\lambda(r)} \\ &\approx \int_{r_0 \in (\rho_0 - \epsilon, \rho_0 + \epsilon)} dr_0 \prod_{j=1}^{\frac{L-1}{2}} \int_{r_j \in [0,\infty) \cap (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j g(r) \delta(C(r)) e^{-\lambda \beta H_\lambda(r)} \\ &\approx \prod_{j=1}^{\frac{L-1}{2}} \int_{r_j \in [0,\infty) \cap (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j g(F_\rho(\bar{r})) e^{-\lambda \beta H_\lambda(F_\rho(\bar{r}))} \\ &= \prod_{j \in \bar{A}_\rho} \int_{r_j \in [0,\epsilon)} dr_j \prod_{j \in K_L^+ - \bar{A}_\rho} \int_{r_j \in (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j g(F_\rho(\bar{r})) e^{-\lambda \beta H_\lambda(F_\rho(\bar{r}))} \\ &\approx \exp(-\lambda \beta H_\lambda(\rho)) N^{-1}(\lambda L_\rho) \prod_{j \in \bar{A}_\rho} \int_{r_j \in [0,\epsilon)} dr_j \prod_{j \in K_L^+ - \bar{A}_\rho} \int_{r_j \in (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j \times \\ &\quad \times g(F_\rho(\bar{r})) N(\lambda L_\rho) \exp(-\frac{1}{2}(\bar{r} - \bar{\rho})^T \lambda L_\rho (\bar{r} - \bar{\rho})) \end{aligned}$$

Case I:

$$\begin{aligned} I(\lambda) &\approx g(\rho) \exp(-\lambda \beta H_\lambda(\rho)) N^{-1}(\lambda L_\rho) \prod_{j \in K_L^+} \int_{r_j \in (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j \times \\ &\quad \times N(\lambda L_\rho) \exp(-\frac{1}{2}(\bar{r} - \bar{\rho})^T \lambda L_\rho (\bar{r} - \bar{\rho})) \end{aligned}$$

$$\approx g(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho)\prod_{j\in K_L^+}\int_{r_j\in(-\infty,+\infty)}dr_j\times$$
$$\times N(\lambda L_\rho)\exp(-\frac{1}{2}(\bar{r}-\bar{\rho})^T\lambda L_\rho(\bar{r}-\bar{\rho}))$$
$$= g(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho) \tag{25}$$

Case IIA: If $g\in\{g_Z,g_0\}$ or $g=g_k\ k\in K_L^+$ and $k\notin\bar{A}_\rho$ then

$$I(\lambda) \approx (\partial_{\bar{A}_\rho}g)(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho)\prod_{j\in\bar{A}_\rho}\int_{r_j\in[0,\epsilon)}dr_j\prod_{j\in\bar{A}_\rho}r_j\times$$
$$\times\prod_{j\in K_L^+-\bar{A}_\rho}\int_{r_j\in(\rho_j-\epsilon,\rho_j+\epsilon)}dr_jN(\lambda L_\rho)\exp(-\frac{1}{2}(\bar{r}-\bar{\rho})^T\lambda L_\rho(\bar{r}-\bar{\rho}))$$
$$\approx (\partial_{\bar{A}_\rho}g)(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho)\prod_{j\in\bar{A}_\rho}\int_{r_j\in[0,\infty)}dr_j\prod_{j\in\bar{A}_\rho}r_j\times$$
$$\times\prod_{j\in K_L^+-\bar{A}_\rho}\int_{r_j\in(-\infty,+\infty)}dr_jN(\lambda L_\rho)\exp(-\frac{1}{2}(\bar{r}-\bar{\rho})^T\lambda L_\rho(\bar{r}-\bar{\rho}))$$
$$= (\partial_{\bar{A}_\rho}g)(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho)\prod_{j\in\bar{A}_\rho}\int_{r_j\in[0,\infty)}dr_j\times$$
$$\times\prod_{j\in\bar{A}_\rho}r_jN(\lambda L_\rho(\bar{A}_\rho))\exp(-\frac{1}{2}\tilde{r}^T\lambda L_\rho(\bar{A}_\rho)\tilde{r}) \tag{26}$$

where $L_\rho(\bar{A}_\rho)=L_\rho$ with rows $i\notin\bar{A}_\rho$ and columns $j\notin\bar{A}_\rho$ crossed out and $\tilde{r}=\{\bar{r}_j\}_{j\in\bar{A}_\rho}$, $\tilde{\rho}=\{\bar{\rho}_j\}_{j\in\bar{A}_\rho}=0$. (The last transformation reflects the fact that the marginal distributions of jointly Gaussian variables are Gaussian). Similarly, if $g=g_k\ k\in K_L^+$ and $k\in\bar{A}_\rho$ then

$$I(\lambda) \approx \frac{1}{3!}(\partial_{r_k}\partial_{r_k}\partial_{\bar{A}_\rho}g)(\rho)\exp(-\lambda\beta H_\lambda(\rho))N^{-1}(\lambda L_\rho)\prod_{j\in\bar{A}_\rho}\int_{r_j\in[0,\infty)}dr_j\times$$
$$\times r_k^3\prod_{j\in\bar{A}_\rho-\{k\}}r_jN(\lambda L_\rho(\bar{A}_\rho))\exp(-\frac{1}{2}\tilde{r}^T\lambda L_\rho(\bar{A}_\rho)\tilde{r}). \tag{27}$$

Therefore, Case IIA reduces to the evaluation of the zero-mean Gaussian expectations

$$e(\lambda) = E\left[\prod_{j=1}^n r_j\mu(r_j)\right] \text{ and}$$
$$e_k(\lambda) = E\left[r_k^3\mu(r_k)\prod_{\substack{j=1\\j\neq k}}^n r_j\mu(r_j)\right]\quad k\in K_L^+$$

(and eventually the ratios $R(k,\lambda)=\frac{\frac{1}{n}e_k(\lambda)}{e(\lambda)}$) where $\mu(x)=\begin{cases}1 & \text{if } x\geq 0\\ 0 & \text{if } x<0\end{cases}$.

In the general covariance case for $n > 1$ it is very difficult to compute $e(\lambda)$, $e_k(\lambda)$, or $R(k,\lambda) = \frac{\frac{1}{3!}e_k(\lambda)}{e(\lambda)}$. However, $L_\rho(\bar{A}_\rho)$ is diagonal. Specifically

$$(L_\rho)_{k,j} = \left[f_0(\rho_0) + \tau\frac{2}{L}\right]\frac{\frac{4}{L}\rho_j\frac{4}{L}\rho_k}{\left(\frac{2}{L}\rho_0 - 1\right)^2} + \delta_{k,j}\left[f_k(\rho_k) + \tau\frac{4}{L}\right]$$

valid only at the critical point ρ where

$$f_j(r_j) = -2(b_{j,2a} + b_{j,2b}) + 12b_{j,4a}r_j^2$$

defines f_j and τ is the Lagrange multiplier at the critical point. Therefore

$$L_\rho(\bar{A}_\rho) = \mathrm{diag}(f_i(0) + \tau\frac{4}{L}, i \in \bar{A}_\rho),$$

the covariance matrix is $\mathrm{diag}(\lambda^{-1}\left[f_i(0) + \tau\frac{4}{L}\right]^{-1}, i \in \bar{A}_\rho)$, the random variables are actually independent, and the expectations factor.

For a scalar Gaussian zero-mean random variable with standard deviation σ the various moments are

$$Ex^n = \begin{cases} 1 \times 3 \times \cdots \times (n-1)\sigma^n & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases}$$

$$E|x|^n = \begin{cases} 1 \times 3 \times \cdots \times (n-1)\sigma^n & \text{if } n = 2k \\ \sqrt{\frac{2}{\pi}}2^k k!\sigma^{2k+1} & \text{if } n = 2k+1 \end{cases}$$

For any symmetric distribution, $E\left[x^n\mu(x)\right] = \frac{1}{2}E\left[|x|^n\right]$. Therefore,

$$e(\lambda) = \frac{1}{\sqrt{(\lambda 2\pi)^n \prod_{j=1}^n \left[f_j(0) + \tau\frac{4}{L}\right]}} \tag{28}$$

$$e_k(\lambda) = \frac{2}{\sqrt{(\lambda 2\pi)^n \lambda^2 \left[f_k(0) + \tau\frac{4}{L}\right]^2 \prod_{j=1}^n \left[f_j(0) + \tau\frac{4}{L}\right]}} \tag{29}$$

$$= \frac{2}{\lambda\left[f_k(0) + \tau\frac{4}{L}\right]}e(\lambda)$$

$$R(k,\lambda) = \frac{1}{3\lambda\left[f_k(0) + \tau\frac{4}{L}\right]}.$$

This completes Step 11 (eqns. 25, 26, 27, 28, and 29).

The previous calculations are formal rather than rigorous. The difficulty in proving results of this nature lies in (1) the location of the critical point, (2) the restricted region of integration, and (3) the vanishing of g at the critical point. Superficially, the manifold constraint $C(r) = 0$ appears to add a great deal of complexity but this is actually not the case. Specifically, fix $k \in K_L^+$. The equation $C(r) = 0$ can be solved for r_k^2 as a polynomial function of r_0, r_1, ..., r_{k-1}, r_{k+1}, ..., $r_{\frac{L-1}{2}}$. This function can be substituted into $-\beta h_{k,r1}$ leaving an exponent that is polynomial, though now with coupling terms between r_i, r_j.

The problem with the critical point and the region of integration is simply that typically multiple components of the critical point are zero and therefore not only is the critical point on the boundary of the region of integration but the boundary is not smooth at that point. The problem with g vanishing at the critical point is that the numerical coefficients of low order terms in the asymptotic expansion will be zero and, as always, high order terms are extremely difficult to compute explicitly. Furthermore, the specific order of the first nonzero term is data dependent.

12 Computation of $E(\Phi_{r,0}|\{y\})$ and $E(\Phi_k|\{y\})$, and Finally $E(\phi_n|\{y\})$

Results are computed only to first order in the asymptotic parameter λ. Specifically,

$$\begin{aligned} Z(\lambda) &= \int g_Z \delta(C) e^{-\lambda\beta H_\lambda} \\ I_0(\lambda) &= \int g_0 \delta(C) e^{-\lambda\beta H_\lambda} \\ I_k(\lambda) &= \int g_k \delta(C) e^{-\lambda\beta H_\lambda} \quad k \in K_L^+ \end{aligned}$$

are computed to first order in λ with resulting functions $\bar{Z}(\lambda)$, $\bar{I}_0(\lambda)$, and $\bar{I}_k(\lambda)$, and then the desired expectations are computed as the ratios

$$E(\Phi_{r,0}|\{y\}) \approx M_{r,0} = \frac{\bar{I}_0(\lambda)}{\bar{Z}(\lambda)}$$

$$E(\Phi_k|\{y\}) \approx M_k = \frac{\bar{I}_k(\lambda)}{\bar{Z}(\lambda)} \quad k \in K_L^+.$$

Recall that the critical points are the same for all of these integrals. Therefore, assuming that only one dominant critical point denoted ρ needs to be included, there are dramatic simplifications in the ratios.

The first simplification, due to canceling common factors in the ratio, results in

$$M_{r,0} = \begin{cases} \frac{g_0(\rho)}{g_Z(\rho)} & \text{if Case I applies} \\ \frac{(\partial_{\bar{A}_\rho} g_0)(\rho)}{(\partial_{\bar{A}_\rho} g_Z)(\rho)} & \text{if Case IIA applies} \\ \text{complicated} & \text{if Case IIB applies} \end{cases}$$

$$M_k = \begin{cases} \frac{g_k(\rho)}{g_Z(\rho)} & \text{if Case I applies} \\ \frac{(\partial_{\bar{A}_\rho} g_k)(\rho)}{(\partial_{\bar{A}_\rho} g_Z)(\rho)} & \text{if Case IIA applies and } k \notin \bar{A}_\rho \\ R(k,\lambda)\frac{(\partial_{r_k}\partial_{r_k}\partial_{\bar{A}_\rho} g_k)(\rho)}{(\partial_{\bar{A}_\rho} g_Z)(\rho)} & \text{if Case IIA applies and } k \in \bar{A}_\rho \end{cases} \quad k \in K_L^+.$$

Note that these expressions are independent of the asymptotic parameter λ, except for the case with $R(k,\lambda)$.

The second simplification results from the multiplicative structure of g. Specifically, the various derivatives at the critical point ρ are

$$(\partial_{\bar{A}_\rho} g_Z)(\rho) = \left(\prod_{k \in K_L - \bar{A}_\rho} q_k^0(\rho_k)\right)\left(\prod_{k \in \bar{A}_\rho} q_k^{0(1)}(\rho_k)\right)$$

$$(\partial_{\bar{A}_\rho} g_k)(\rho) = \begin{cases} q_k^{1(1)}(\rho_k)\left(\prod_{j \in K_L - \bar{A}_\rho} q_j^0(\rho_j)\right)\left(\prod_{j \in \bar{A}_\rho - \{k\}} q_j^{0(1)}(\rho_j)\right) & \text{if } k \in \bar{A}_\rho \\ q_k^1(\rho_k)\left(\prod_{j \in K_L - \{k\} - \bar{A}_\rho} q_j^0(\rho_j)\right)\left(\prod_{j \in \bar{A}_\rho} q_j^{0(1)}(\rho_j)\right) & \text{if } k \notin \bar{A}_\rho \end{cases} \quad k \in K_L$$

$$(\partial_{r_k}\partial_{r_k}\partial_{\bar{A}_\rho} g_k)(\rho) = \begin{cases} q_k^{1(3)}(\rho_k)\left(\prod_{j \in K_L - \bar{A}_\rho} q_j^0(\rho_j)\right)\left(\prod_{j \in \bar{A}_\rho - \{k\}} q_j^{0(1)}(\rho_j)\right) & \text{if } k \in \bar{A}_\rho \\ q_k^{1(2)}(\rho_k)\left(\prod_{j \in K_L - \{k\} - \bar{A}_\rho} q_j^0(\rho_j)\right)\left(\prod_{j \in \bar{A}_\rho} q_j^{0(1)}(\rho_j)\right) & \text{if } k \notin \bar{A}_\rho \end{cases} \quad k \in K_L.$$

Therefore,

$$M_{r,0} = \begin{cases} \rho_0 & \text{if Cases I or IIA} \\ \text{complicated} & \text{if Case IIB} \end{cases}$$

$$M_k = \begin{cases} \frac{\Psi_k^*}{|\Psi_k|}\rho_k\frac{\Theta_k^1(\rho_k)}{\Theta_k^0(\rho_k)} & \text{if Case I applies} \\ \frac{\Psi_k^*}{|\Psi_k|}\rho_k\frac{\Theta_k^1(\rho_k)}{\Theta_k^0(\rho_k)} & \text{if Case IIA applies and } k \notin \bar{A}_\rho \\ R(k,\lambda)\frac{q_k^{1(3)}(\rho_k)}{q_k^{0(1)}(\rho_k)} & \text{if Case IIA applies and } k \in \bar{A}_\rho \end{cases}.$$

It is now possible to combine cases and give a simpler characterization of when the cases apply because each of the M_k depends only on ρ_k. Specifically,

$$M_{r,0} = \begin{cases} \rho_0 & \text{if } \rho_0 \neq 0 \\ \text{complicated} & \text{if } \rho_0 = 0 \end{cases} \tag{30}$$

$$M_k = \begin{cases} \frac{\Psi_k^*}{|\Psi_k|}\rho_k \frac{\Theta_k^1(\rho_k)}{\Theta_k^0(\rho_k)} & \text{if } \rho_k \neq 0 \\ R(k,\lambda)\frac{q_k^{1(3)}(\rho_k)}{q_k^{0(1)}(\rho_k)} & \text{if } \rho_k = 0 \end{cases} \quad k \in K_L^+. \tag{31}$$

Finally, recall from Section 7 that $E(\Phi_k|\{y\}) = E(\Phi_{L-k}|\{y\})^*$ and therefore set $M_k = M_{L-k}^*$ $k \in \{\frac{L-1}{2}+1, \ldots, L-1\}$. This completes Step 12.

The M_k are the approximations to $E(\Phi_k|\{y\})$. Therefore, the approximation to the optimal estimate of the field ϕ_n is completed with two steps. First, compute an approximation, denoted m_n, to $E(\phi_n|\{y\})$ by computing the inverse Fourier series of the M_k. Second threshold m_n at $\frac{1}{2}$ to compute the final estimate $\hat{\phi}_n$ which is the reconstructed signal. Specifically,

$$\hat{\phi}_n = \begin{cases} 1 & \text{if } m_n \geq \frac{1}{2} \\ 0 & \text{if } m_n < \frac{1}{2} \end{cases} . \tag{32}$$

Sites n where $\hat{\phi}_n$ is 1 are occupied by a generic atom while the remaining sites are unoccupied. Finally, phase angle estimates, if desired, can be computed as the phase of the Fourier series coefficients of the thresholded field $\hat{\phi}_n$. This completes Step 13.

Note that reference to Problem 1 versus Problem 2 asymptotics does not occur in the solution. Rather, as discussed in Section 9, the choice is hidden in the definitions of the constants $b_{k,ns}$ defined in Appendix A.

In summary, the signal reconstruction algorithm has the following steps.

(1) Choose Problem 1 versus Problem 2 asymptotics and compute the appropriate $b_{k,ns}$ (Appendix A).

(2) Compute the critical point (eqns. 23, 24, 19, and 21; see also [13]).

(3) Compute M_k (eqns. 30 and 31).

(4) Compute m_n, the inverse Fourier series of M_k.

(5) Compute the reconstructed signal $\hat{\phi}_n$ by thresholding m_n (eqn. 32).

13 Discussion and Future Directions

In this paper I have proposed and (approximately) solved a Bayesian signal-reconstruction problem motivated by an X-ray crystallography inverse problem. The unusual part of the Bayesian model from the image phase-retrieval perspective is the 0-1 nature of the object and the presence of an a priori density described by a MRF. The unusual part of the model from the crystallographic perspective is the absence of explicit scattering phase variables, the detailed MRF-based a priori model for atomic locations, and the detailed modeling of observation errors. The unusual aspects of the solution are the concern with symmetry breaking, the use of the spherical model for a fixed-size lattice, and the asymptotic approximation in terms of small observation-noise variances.

The resulting estimator has an interesting structure. Fix the kernel ψ of $H^{\text{s.b.}}$ (eqn. 1). The location of the critical point (eqns. 23, 24, 19, and 21) is independent of ψ. However, $\angle E(\Phi_k|\{y\}) = -\angle\Psi_k$ (eqns. 30 and 31) and furthermore g (eqns. 13, 14, 11, 12, and 10) and therefore $|E(\Phi_k|\{y\})|$ (eqn. 31) depends on ψ. In the final form for $|E(\Phi_k|\{y\})|$ given in eqn. 31 the dependence on ψ enters through the dependence of Θ_k^0 and Θ_k^1 on $|\Psi_k|$ (eqns. 5 and 9).

The fact that the angle of the estimate is exactly $-\angle\Psi_k$ is reminiscent of Fienup-Gerchberg-Saxton type algorithms [21] where the phase function is constant around an iteration until the Object space update step. However, the present situation is quite different because ψ is the kernel of the symmetry breaking function and because this is not an iterative algorithm–for fixed ψ one makes only one transformation from Fourier to Object space. However, the appropriate choice of ψ for a given problem is not clear. In [13] one iterative method is discussed. However, the resemblance to Fienup-Gerchberg-Saxton type algorithms might lead to a better method.

Again note that any pattern of missing observations and any pattern of k-dependent observation noise variance are admissible. Furthermore, the calculations have essentially no

dependence on the dimension d of the space. Finally, the calculation, for fixed ψ, is very quick. The computation of the critical point and evaluation of the asymptotic formulae is linear in the number of sites in the lattice independent of the dimension d so the order of the total calculational burden is dominated by the calculation of the inverse Fourier series from M_k to m_n, the approximate conditional mean of ϕ_n.

Finally, in [13], I describe data-adaptive ideas for the choice of ψ, define parameters in H^{apriori} that are suitable for modeling bond-length limitations in covalently-bound molecules, and give several numerical examples in one and two dimensions on simulated data. The examples include a tiny one-dimensional problem where it is possible to compute the estimator performance statistics versus observation noise intensity for the exact conditional mean estimator by brute force and compare with the approximate estimators.

In the future, improvement of the data adaptive ideas described in [13] and introduction of general space groups into these calculations are important goals. Though not necessary for an imaging application, the latter is necessary before crystallographic data can be processed.

14 Acknowledgements

The use of ideas from statistical mechanics (e.g., the spherical model) and Markov random fields grew out of discussions with Professor Sanjoy K. Mitter while the author was a Postdoctoral Associate at the Laboratory for Information and Decision Systems at M.I.T., Cambridge, MA. I would like to thank Professors Sanjoy K. Mitter and Alan S. Willsky for technical interest and financial support, Professor Bernard Gaveau for important observations on the asymptotic calculations, and Dr. Charles Rockland for mathematical advice. A detailed reading by Professor Willsky greatly improved the manuscript. This work was supported at M.I.T. by the Air Force Office of Scientific Research (AFOSR-89-0276), the Army Research Office (DAAL03-86-K-0171), and the National Science Foundation (ECS-8700903) and at Purdue University by a Whirlpool Faculty Fellowship and the School of Electrical

Engineering, Purdue University.

15 Appendix A

This appendix defines the $a_{k,j}$ constants (see Section 6) and the $b_{k,ns}$ constants (see Sections 8 and 9). The $a_{k,j}$ are

$$a_{k,0} = \begin{cases} \frac{-\beta}{2\sigma_0^2} y_0^2 & \text{if } k = 0 \\ \frac{-\beta}{2\sigma_k^2} y_k^2 + \frac{-\beta}{2\sigma_{L-k}^2} y_{L-k}^2 & \text{if } k \neq 0 \end{cases}$$

$$a_{k,1} = \begin{cases} -\beta w_1 - \beta q \Psi_0 & \text{if } k = 0 \\ -\beta 2q|\Psi_k| & \text{if } k \neq 0 \end{cases}$$

$$a_{k,2} = \begin{cases} -\frac{\beta}{L} W_2(0,0) + \frac{\beta}{\sigma_0^2} y_0 & \text{if } k = 0 \\ -\frac{2\beta}{L} W_2(-k,k) + \frac{\beta}{\sigma_k^2} y_k + \frac{\beta}{\sigma_{L-k}^2} y_{L-k} & \text{if } k \neq 0 \end{cases}$$

$$a_{k,4} = \begin{cases} \frac{-\beta}{2\sigma_0^2} & \text{if } k = 0 \\ \frac{-\beta}{2\sigma_k^2} + \frac{-\beta}{2\sigma_{L-k}^2} & \text{if } k \neq 0 \end{cases}$$

There are two different sets of $b_{k,ns}$ constants corresponding to the two different definitions of the asymptotics. For Problem 1 the definitions are

$$b_{k,0a} = \begin{cases} \frac{\beta}{2\bar{\sigma}_0^2} y_0^2 & \text{if } k = 0 \\ \frac{\beta}{2\bar{\sigma}_k^2} y_k^2 + \frac{\beta}{2\bar{\sigma}_{L-k}^2} y_{L-k}^2 & \text{if } k \neq 0 \end{cases}$$

$$b_{k,1c} = \begin{cases} -\beta w_1 - \beta q \Psi_0 & \text{if } k = 0 \\ -\beta 2q|\Psi_k| & \text{if } k \neq 0 \end{cases}$$

$$b_{k,2a} = \begin{cases} \frac{\beta}{\bar{\sigma}_0^2} y_0 & \text{if } k = 0 \\ \frac{\beta}{\bar{\sigma}_k^2} y_k + \frac{\beta}{\bar{\sigma}_{L-k}^2} y_{L-k} & \text{if } k \neq 0 \end{cases}$$

$$b_{k,2c} = \begin{cases} -\frac{\beta}{L} W_2(0,0) & \text{if } k = 0 \\ -\frac{2\beta}{L} W_2(-k,k) & \text{if } k \neq 0 \end{cases}$$

$$b_{k,4a} = \begin{cases} \frac{\beta}{2\bar{\sigma}_0^2} & \text{if } k = 0 \\ \frac{\beta}{2\bar{\sigma}_k^2} + \frac{\beta}{2\bar{\sigma}_{L-k}^2} & \text{if } k \neq 0 \end{cases}$$

For Problem 2 the definitions are

$$b_{k,0a} = \begin{cases} \frac{\beta}{2\bar{\sigma}_0^2} y_0^2 & \text{if } k = 0 \\ \frac{\beta}{2\bar{\sigma}_k^2} y_k^2 + \frac{\beta}{2\bar{\sigma}_{L-k}^2} y_{L-k}^2 & \text{if } k \neq 0 \end{cases}$$

$$
\begin{aligned}
b_{k,1b} &= \begin{cases} -\beta\chi w_1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases} \\
b_{k,1c} &= \begin{cases} -\beta q \Psi_0 & \text{if } k = 0 \\ -\beta 2q|\Psi_k| & \text{if } k \neq 0 \end{cases} \\
b_{k,2a} &= \begin{cases} \frac{\beta}{\bar{\sigma}_0^2} y_0 & \text{if } k = 0 \\ \frac{\beta}{\bar{\sigma}_k^2} y_k + \frac{\beta}{\bar{\sigma}_{L-k}^2} y_{L-k} & \text{if } k \neq 0 \end{cases} \\
b_{k,2b} &= \begin{cases} -\frac{\beta}{L}\chi W_2(0,0) & \text{if } k = 0 \\ -\frac{2\beta}{L}\chi W_2(-k,k) & \text{if } k \neq 0 \end{cases} \\
b_{k,4a} &= \begin{cases} \frac{\beta}{2\bar{\sigma}_0^2} & \text{if } k = 0 \\ \frac{\beta}{2\bar{\sigma}_k^2} + \frac{\beta}{2\bar{\sigma}_{L-k}^2} & \text{if } k \neq 0 \end{cases}
\end{aligned}
$$

References

[1] Herbert Hauptman. Direct methods and anomalous dispersion. *Chemica Scripta*. 26:277–286, 1986.

[2] Jerome Karle. Recovering phase information from intensity data. *Chemica Scripta*. 26:261–276, 1986.

[3] M. M. Wolfson. *An Introduction to X-ray Crystallography.* Cambridge University Press, London, 1970.

[4] Carmelo Giacovazzo. *Direct Methods in Crystallography.* Academic Press, London, 1980.

[5] Herbert A. Hauptman. *Crystal Structure Determination: The Role of the Cosine Semi-invariants.* Plenum Press, New York, 1972.

[6] G. Bricogne. Maximum entropy and the foundations of direct methods. *Acta Crystallographica A*, 40:410–445, 1984.

[7] G. Bricogne. A Bayesian statistical theory of the phase problem. I. Multichannel maximum entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta Crystallographica A*, 44:517–545, 1988.

[8] R. P. Millane. Phase retrieval in crystallography and optics. *Journal of the Optical Society of America A*, 7(3):394–411, 1990.

[9] Henry Stark, editor. *Image Recovery: Theory and Application.* Academic Press, Orlando, 1987.

[10] T. H. Berlin and M. Kac. The spherical model of a ferromagnet. *The Physical Review*. 86(6):821–835, 1952.

[11] Rodney J. Baxter. *Exactly Solved Models in Statistical Mechanics.* Academic Press. London, 1982.

[12] Ross Kindermann and J. Laurie Snell. *Markov Random Fields and Their Applications.* American Mathematical Society, Providence, Rhode Island, 1980.

[13] Peter C. Doerschuk. Adaptive Bayesian signal reconstruction with a priori model implementation and synthetic examples for x-ray crystallography. *Journal of the Optical Society of America A*, 1991. To appear.

[14] Jose Luis Marroquin. *Probabilistic Solution of Inverse Problems.* PhD thesis, M.I.T., Cambridge, MA 02139, September 1985.

[15] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association (Theory and Methods)*, 82(397):76–89, 1987.

[16] Alfred Levitas and Melvin Lax. Statistics of the Ising ferromagnet. *Physical Review*, 110(5):1016–1027, June 1958.

[17] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, New York, 4 edition, 1980.

[18] Carl M. Bender and Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill Book Company, New York, 1978.

[19] Dimitri P. Bertsekas. Notes on nonlinear programming and discrete-time optimal control. Technical Report LIDS-R-919, Laboratory for Information and Decision Systems, MIT, Cambridge MA 02139, July 1979.

[20] P. P. Varaiya. *Notes on Optimization.* Van Nostrand Reinhold Company, New York, 1972.

[21] J. R. Fienup. Phase retrieval algorithms: A comparison. *Appl. Opt.*, 21:2758–2769, 1982.