4-23-1992

Bayesian Signal Reconstruction from Fourier Transform Magnitude in the Presence of Symmetries and X-ray Crystallography

Peter C. Doerschuk
Purdue University School of Electrical Engineering

BAYESIAN SIGNAL RECONSTRUCTION FROM FOURIER TRANSFORM MAGNITUDE IN THE PRESENCE OF SYMMETRIES AND X-RAY CRYSTALLOGRAPHY

PETER C. DOERSCHUK

TR-EE 92-16
APRIL 1992

SCHOOL OF ELECTRICAL ENGINEERING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1285

Bayesian Signal Reconstruction from. Fourier Transform Magnitude in the Presence of Symmetries and X-ray Crystallography

Peter C. Doerschuk

School of Electrical Engineering

Purdue University

West Lafayette, Indiana 47907–1285

(317) 494–1742

(317) 494–6440 (FAX)

doerschu@ecn.purdue.edu

April 23, 1992

Abstract

In Ref. [1] a signal reconstruction problem motivated by x-ray crystallography was solved using 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. In Ref. [2] a detailed parameterization of the a *priori* model appropriate for crystallography was proposed and symmetry-breaking parameters in the solution were

usecl to perform data-dependent adaptation of the estimator. The adaptation attempts to minimize the effects of the spherical model approximation used in the solution. In this paper these ideas are extended to signals that obey a space group symmetry, which is a crucial extension for the x-ray crystallography application. Performance statistics for reconstruction in the presence of a space group symmetry based on simulated data are presented.

[1] Peter C. Doerschuk. "Bayesian Signal Reconstruction, Markov Random Fields, and X-Ray Crystallography." *Journal* of the Optical Society of America A, 8(8):1207–1221, 1991.

[2] 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, 8(8):1222–1232, 1991.

1 Introduction

In Ref. [1] a novel Bayesian statistical approach was presented to a class of phase-retrieval problems exemplified by the inverse problem of single-crystal x-ray crystallography. In Ref. [2] parameters were proposed in the a *priori* model that are suitable for the x-ray crystallography application, free parameters in the estimator were employed in order to design a data-dependent adaptive estimator, and several numerical examples were presented. In this paper these ideas are extended to signals that are invariant under the actions of a space group symmetry. This extension is crucial for the x-ray crystallography application since essentially all crystallographic data displays a space group symmetry. Similar symmetries occur in computer vision problems. Three different approaches to the extension are presented and, for a particular space group, the three algorithms that result are compared numerically on simulated data.

The novel contribution of this paper is the incorporation of space group symmetries [3, 4, 5]. A space group, denoted $\mathcal{G}$, is a set of operators on vectors in R^d where the operators form a group. To say that a function $p : \mathrm{R}^d \rightarrow \mathrm{R}$ is invariant under the actions of the space group means that $p(Tx) = p(x)$ for any $x \in \mathrm{R}^d$ and any $\mathrm{T} \in \mathcal{G}$. The theory of space groups deals with topics such as the number of space groups for a given dimension d, their subgroup relationships, methods of describing the operators T in the group, and so forth.

The purpose of an x-ray crystallography experiment is to measure the positions in three dimensional space of each atom making up a molecule of interest. The data is the magnitude of the Fourier transform of the electron density in a crystal composed of the molecule of interest. Because the electron density is highly peaked around the nuclear locations, the desired three dimensional locations can be found by reconstructing the electron density. Hence this is a phase retrieval problem. Millane [6] summarizes and compares phase retrieval methods in optics and crystallography. An important contrast is that in x-ray crystallography the Fourier transform magnitude is sampled, the sampling is due to the periodic crystal

structure and is therefore fixed, and the sampling is an undersampling. Therefore, uniqueness of the solution is a serious issue which, however, is not addressed in this paper. In the x-ray crystallography application it is the electron density in the crystal that is invariant under the space group. Furthermore, the identity of the space group is determined by preliminary experiments and can be considered known before the electron density reconstruction is done. Therefore the x-ray crystallography problem is to reconstruct a signal that is known, among other information, to be invariant under a particular space group symmetry.

In any dimension d there is a trivial space group $P1$ with d operators which are translation by the period (possibly different) in each of the d directions. This is exactly the idea of a periodic signal in d dimensions. Therefore, though it was not emphasized, the work in Refs. [1, 2] actually concerned signals invariant under the trivial space group. In crystallography terminology the d-dimensional repeat unit of the periodic signal is called the unit cell. The change in going from the trivial space group to any nontrivial space group is that in the nontrivial space group there is structure within the unit cell. An example of such structure is division of the unit cell in half with one half the mirror image of the other half.

The approach taken to the x-ray crystallography problem in Refs. [1, 2] and the current paper is Bayesian signal reconstruction in the spirit of, for example, Ref. [7, 8, 9]. There is a periodic object which takes only values zero and one, and the observerr makes corrupted measurements, denoted y_k, of the magnitude of the object's Fourier transform. The goal is to reconstruct the object from these measurements and from a *priori* probabilistic information concerning the class of likely objects. One period of the object is modeled as a binary-valued finite-lattice Markov random field (MRF) denoted ϕ_n and the corruption of the measurements is modeled as additive independent zero-mean Gaussian random variables with k-dependent variances. Both the MRF and the Gaussian observittion errors can be put in the form of energy functions with the corresponding probabilities in the form of the Gibbs distribution. A Bayesian estimation problem is approximately solved. The criteria is to minimize the mean squared reconstruction error for the field ϕ. [Throughout this paper,

a subscripted variable (i.e., "ϕ_n") appearing without subscript (i.e., "ϕ") means the set of variables as the subscript ranges over its possible values]. Therefore the optimal estimator first computes the conditional mean of the field ϕ given the data y, i.e., $E(\phi_n|y)$, and then thresholds the conditional mean at value 1/2 to derive the optimal estimate $\hat{\phi}_n$, i.e., $\hat{\phi}_n = \begin{cases} 1, & E(\phi_n|y) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$. In order to compute $E(\phi_n|y)$ one computes averages. However, because only the magnitude of the Fourier transform is available, the definition of the coordinate system on the object is lost. For example, in one dimension, all information concerning the origin in the sense of translation and concerning the handedness in the sense of inversion through the origin ($n \rightarrow -n$) is lost. Furthermore, the a *priori* information described by the MRF is also invariant under translation and inversion. If one blindly averages over all configurations of ϕ, the result is a DC field. Therefore, an additional term in the energy function of the MRF is introduced which favors certain configurations. For example, in one dimension, the additional term breaks the symmetries of the previous energy function with respect to translation and inversion. The term has the form of a convolution of the field ϕ_n with a kernel function ψ_n.

There are several differences between this approach and traditional methods in x-ray crystallography. The first difference is 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 the approach proposed in these papers is to invest a great deal of effort in modeling of the correlations between the atomic locations. That is, a large effort is made to improve the

accuracy of the the chemistry model. At present, 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. The approach proposed in these papers includes these inaccuracies in a fundamental way from the very start of the calculation.

In solving the estimation problem in Ref. [1] two approximations were made. The first approximation was the spherical model which relaxed the 0-1 nature of the lattice variables ϕ_n. The second approximation was to evaluate integrals asymptotically as the observation noise variance approached zero. Given these two approximations, explicit (e.g., no numerical quadratures or nonlinear optimizations) formulae were computed for an approximation, denoted m_n, to $E(\phi_n|y)$ as a function of ψ and y. These formulae are easy to compute, can accommodate missing data and varying observation noise variance, and are essentially the same in any dimensional space. Finally, the estimate $\hat{\phi}_n$ of the field ϕ_n is m_n thresholded at 1/2: $\hat{\phi}_n = \begin{cases} 1, & m_n \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$. The emphasis on analytical calculations contrasts with much of the estimation work based on MRFs which is either simulation oriented [7, 8, 9] or requires restrictions on the neighborhood structure of the MRF [10].

For any choice of ψ this is a valid Bayesian estimation problem. Therefore, the ψ are chosen by optimizing a data-dependent cost function that minimizes the effects of the two approximations made in the solution of the estimation problem.

In this paragraph the Hamiltonian is recalled from Ref. [1]. The Harniltonian is in three parts–the a *priori* probability part $H^{apriori}$, the conditional observational probability part H^{obs}, arid the symmetry breaking part $H^{s.b.}$. That is, $H = H^{apriori} + H^{obs} + H^{s.b.}$. Equations are stated for the one-dimensional case. In d dimensions exactly the same equations hold with indices, lattice dimensions, and sums all expanded to d dimensions. The a *priori* part

is the most general shift-invariant quadratic, specifically,

$$u_n = \sum_{n_1=0}^{L-1} \sum_{n_2=0}^{L-I} \phi_{n+n_1} w_2(n_1, n_2) \phi_{n+n_2} + w_1 \phi_n \tag{1}$$

$$H^{\text{apriori}} = \sum_{n=0}^{L-1} u_n,$$

where L is the size of the lattice which is also the period of the crystal when measured in lattice spacings. The conditional observational part is Gaussian, specifically,

$$z_k = |\Phi_k|^2$$

$$\Phi_k = \sum_{n=0}^{L-1} \phi_n e^{-jnk\frac{2\pi}{L}}$$

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

where y_k and σ_k are observed in the experiment. However, the σ_k values are assumed to be exact. Finally, the symmetry breaking part is a convolution of the field ϕ with the kernel function ψ, specifically,

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

where ψ_n is real and periodic with period L.

A major concern in Ref. [1, 2] was symmetry breaking. Recall that symmetry breaking was necessary because all information about the origin and the handedness of the coordinate system for the electron density is lost when the phases of the Fourier coefficients of the electron density are not recorded. For example, in one dimension, $x(t)$, $x(t+\tau)$, and $x(-t)$ have Fourier transforms with the same magnitude function. However, when the signal is known to satisfy a particular space group symmetry, then there is partial or full retention of this information even though the phases of the Fourier coefficients are still not recorded. Therefore it may be possible to do away with symmetry breaking. Dropping symmetry breaking has both positive and negative aspects. It is desirable since symmetry breaking does not have a basis in the physical model and since it requires a numerical optimization of the kernel of the symmetry breaking function. It is undesirable because the data adaptation, which attempts

to minimize the effects of the spherical model approximation, occurs through optimization of the symmetry breaking function. One of the major conclusions of the numerical experiments reported here is that data adaptation is important and therefore retaining symmetry breaking is desirable.

From the point of view of signal reconstruction, ignoring the space group causes two problems.

1. The presence of a space group implies that the electron density is equal at physically separated locations. This. reduces the number of electron density variables that have to be estimated from the data. Estimation of fewer variables from the same data set improves the quality of each estimate. In three dimensional space groups, the reduction is usually by a factor of four or more. If the presence of the space group is ignored, then it is not possible to directly exploit this reduction.

2. If the space group is ignored, then the estimated electron density. will typically not exhibit the space group symmetry because of inaccuracies in the data.

Furthermore, if the space group is ignored, then symmetry breaking is necessary.

There are three approaches to solving signal reconstruction problems in the presence of nontrivial space groups using extensions of the ideas in Ref. [1, 2]. In Approach 1, the basic point of view is to replace the space group $\mathcal{G}$ present in the data by the subgroup $P1$. The resulting;signal reconstruction problem has been solved [1, 2]. Then the space group information is added back into the signal reconstruction problem in two ways: First, reconstructions that are invariant under $P1$ but not $\mathcal{G}$ are transformed into reconstructions invariant under $\mathcal{G}$ by averaging. This solves Problem 2 and ameliorates Problem 1. Second, the invariance of the signal under $\mathcal{G}$ is applied as a soft constraint by adding a term to the symmetry breaking;optimization criteria. This ameliorates Problem 1. The advantage of Approach 1 is simplicity since the work of Ref. [1, 2] is applied with little alteration to any space group $\mathcal{G}$. The disadvantage is the suboptimal use of space group information. Symmetry breaking is

retained.

The second and third approaches both integrate the presence of the space group $\mathcal{G}$ as a hard constraint into the signal reconstruction process rather than treating it as primarily a soft constraint added to fundamentally $P1$ oriented processing. The two approaches differ by the order in which noncommuting nonlinear operations are performed: in Approach 2 the spherical model is applied before the space group symmetry is enforced while in Approach 3 the order is reversed. In both cases the symmetry constraint is applied as a hard constraint that is satisfied exactly. The advantage of Approach 2 is that the calculation of the critical point in the small observation noise asymptotics is essentially unchanged from the corresponding (calculationin Refs. [1][2, Appendix A]. Therefore it can be done analytically. The disadvantage is that the spherical model approximation is applied over a larger number of sites (the entire unit cell) and so it is less accurate. Symmetry breaking is required. The advantage of Approach 3 is that the spherical model is applied over a smaller number of sites (only the fundamental domain) and so it is more accurate. The disadvantage is that the calculation of the critical point in the small observation noise asymptotics is substantially more difficult than the corresponding calculation in Refs. [1][2, Appendix A] and to date an analytical solution is available only for a special case. Symmetry breaking is not required, mirroring the fact that symmetry breaking is not required in an exact solution. In fact, if used, symmetry breaking only influences the value of second and higher order terms in the asymptotic expansion.

The purpose of this paper is to present methodology for the extension of ideas in Ref. [1, 2] to signal reconstruction problems where the signal is invariant under the actions of a space group. Because the calculations are complicated in two and three dimensions, the methods are illustrated in one dimension. The desire for a one dimensional example chooses the space group because in one dimension there are only two space groups [4, p. 12]: the group $P1$ treated in Ref. [1, 2] and the group $P\bar{1}$ used as the example in this paper. These calculations are presently being continued, with surprisingly little modification, for the monoclinic C2

space group in three dimensions. This space group was chosen because it is not centrosymmetric (so the Fourier coefficients of the electron density will be complex) and because I happened to be given data in this space group. It is convenient that monoclinic C2 is a symmorphic [3, p. 114] space group.

The remainder of this paper is organized in the following fashion. In Section 2 the $P\bar{1}$ space group is described. The numerical example that is studied throughout this paper is introduced in Section 3. Approach 1 based on averaging and soft constraints is described in Section 4. Numerical results for Approach 1 are discussed in Section 5. In the second half of the paper Approaches 2 and 3 are presented jointly through a sequence of sections starting with an introduction (Section 6). In Section 7 the spherical model is covered. For Approach 3, where tbe symmetry constraint is applied before the spherical model, the constraint is also covered. In Section 8 the Fourier coordinates are discussed. For Approach 2, where the symmetry constraint is applied after the spherical model, the constraint is also discussed. Approaches 2 and 3 are parallel for Sections 9-11. In these sections the Bayesian integrals in Fourier coordinates (Section 9), the motivation and definition of the small noise asymptotics (Section 10), and the notation for the asymptotic evaluation of the Bayesian integrals (Section 11) are presented. The presentation then diverges with sections on Approach 2 [calculation of the critical point (Section 12), asymptotic formulae (Section 13), and numerical results (Section 14)] and a corresponding development concerning Approach 3 [calculation of the critical point (Section 15), asymptotic formulae (Section 16), and numerical results (Section 17)]. Finally, the results to date and direction for future research are discussed in Section 18.

2 The $P\bar{1}$ Space Group

In this section the Pi space group is described and several properties are noted.

Space groups are typically studied as transformations on R^d. However, because this

paper applies these ideas to MRF lattices, space groups are described as transformations on Z^d. For many space groups, including the $P\bar{1}$ space group of interest in this paper and the three-dimensional monoclinic C2 space group for which crystallographic data is available, there are simple discretizations of R^d which preserve the necessary properties of the space group. There may exist space groups for which this is a more difficult transformation.

Part of the definition of a space group is the dimension d of the space. For that reason it is strictly speaking incorrect to discuss $P\bar{1}$ without giving a dimension. The standard notation, which includes the dimension in the name, is $p\bar{1}$ for d = 1 [4, p. 38], $p2$ for d = 2 [4, p. 83], and $P\bar{1}$ for d = 3 [4, p. 104]. As is seen in the following, these three groups are so close in concept that in this paper the same label $P\bar{1}$ is used for all of them. The same comments apply to what is called the $P1$ space group in Section 1. In that case the standard notation is p1 for d = 1 [4, p. 38], p1 for d = 2 [4, p. 82], and $P1$ for d = 3 [4, p. 102].

In one dimension, signals invariant under the $P\bar{1}$ space group are periodic signals that are symmetric around the midpoint of the period. In more detail, let the signal be ϕ. Then ϕ is invariant under $P\bar{1}$ if there exists an L such that $\phi_n = \phi_{n+L}$ (translation by L) and $\phi_n = \phi_{-n}$ (inversion through the origin). L is the period and, since the signal is periodic, the inversion condition can be rewritten as $\phi_n = \phi_{L-n}$ which makes clearer the symmetry around the midpoint of the period.

In the crystallography application, the preliminary experiments mentioned in Section 1 provide both the space group, i.e., Pi, and the value of the group parameters, i.e., L for $P\bar{1}$ in one dimension. In other words, L does not have to be estimated in the course of the reconstruction.

In d dimensions $P\bar{1}$ has several instead of just one translation operation. Let $\vec{e}_i$ be the i th standard unit vector in R^d. Then ϕ is invariant under Pi in d dimensions if there exist $L_1, \ldots, L_d$ such that for each i $\in (1, \ldots, d)$, $\phi_{\vec{n}} = \phi_{\vec{n}+L_i\vec{e}_i}$ (translation by L_i in coordinate i), and in addition $\phi_{\vec{n}} = \phi_{-\vec{n}}$ (inversion through the origin).

Note that in any dimension the Fourier coefficients of a signal invariant under $P\bar{1}$ are real

since $\phi_{\vec{n}} = \phi_{-\vec{n}}$. This simplification likely contributes to the performance of the algorithms described here. However, related algorithms have already been shown to work in the presence of $P1$ symmetry in which case the Fourier coefficients are complex.

As described in Section 1, the d-dimensional repeat unit of the periodic signal is called the unit cell. It is not unique. For the one-dimensional $P\bar{1}$ space group it is most convenient to take the unit cell as $U = \{0, \ldots, L-1\}$.

The fundamental domain of a space group is the smallest region $F \subset Z^d$ such that knowledge of $\phi_{\vec{n}}$ for $\vec{n} \in F$ implies knowledge of $\phi_{\vec{n}}$ for all $\vec{n} \in Z^d$. It is also not unique. For the one-dimensional $P\bar{1}$ space group it is most convenient to take

$$F = \begin{cases} \{0, \ldots, \frac{L-1}{2}\}, & L \text{ odd} \\ \{0, \ldots, \frac{L}{2}\}, & L \text{ even} \end{cases}$$

That the preceding choice of F is adequate can be verified in two steps: first use inversion ($\phi_n = \phi_{-n}$) to compute ϕ_n on an entire period and then use translation ($\phi_n = \phi_{n+L}$) to extend the period to all of Z.

An orbit [4, p. 724] of a space group is the set of all locations in Z^d that can be reached from a particular location in Z^d by application of the space group operations. If N is an orbit of the space group and ϕ is invariant under the space group then $\phi_{\vec{n}_1} = \phi_{\vec{n}_2}$ for all $\vec{n}_1, \vec{n}_2 \in N$. It may well be that $\phi_{\vec{n}}$ takes this same value for some $\vec{n} \notin N$, but that is purely fortuitous. Clearly exactly one point of each orbit must fall in the fundamental domain. An orbit defined in this fashion is typically infinite in size because of the translation operator. For the one-dimensional $P\bar{1}$ case with L odd there are $\frac{L-1}{2} + 1$ orbits which are $\{nL : n \in Z\}$ and $\{m+nL : n \in Z\} \cup \{-m+nL : n \in Z\}$ form $m \in \{1, \ldots, \frac{L-1}{2}\}$ while for L even there are $\frac{L}{2}+1$ orbits which are $\{nL : n \in Z\}$, $\{\frac{L}{2}+nL : n \in Z\}$, and $\{m+nL : n \in Z\} \cup \{-m+nL : n \in Z\}$ for $m \in 1, \ldots, -1$ In the following, the term 'orbit' is used to mean the intersection of these infinite sets with the unit cell U. These finite sets are, for L ocld, $\{0\}$, $\{1, L-1\}$, $\ldots$, $\{\frac{L-1}{2}, \frac{L+1}{2}\}$ and, for L even, $\{0\}$, $\{1, L-1\}$, $\ldots$, $\{-1, \frac{L}{2}+1\}$, $\{\frac{L}{2}\}$. The length of an orbit is the number of elements in the set. For the one-dimensional $P\bar{1}$ space group the

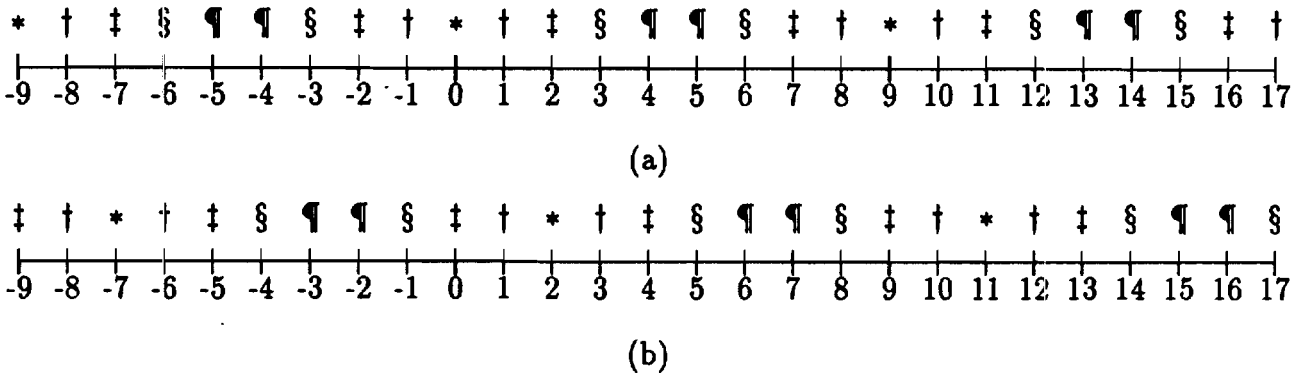


Figure 1: Symmetry breaking in $P\bar{1}$ for L odd, specifically L = 9: different symbols represent values that need not be equal. (a) Example of an invariant signal. (b) Example translation, specifically by 2, of the signal from (a) demonstrating the lack of invariance in the translated signal.

length is either 1 or 2.

As described in Section 1, knowledge that a signal is invariant under the actions of a particular space group can obliterate the need for symmetry breaking. Consider a signal ϕ in one dimension that is invariant under $P\bar{1}$ with L odd. In that case L − 1 of the points in the unit cell are paired in orbits of length 2 therefore sharing the same value and one point in the unit cell is isolated in an orbit of length 1 therefore having a unique value. First consider the choice of origin location (Figure 1). If ϕ' is some translation of ϕ by an interval not equal to the period then ϕ' will typically not be invariant. For instance, the point isolated in an orbit of length 1 and therefore having a unique value will now be in an orbit of length 2 and the other point in the orbit will have a different value. Therefore, even though the phase of the Fourier coefficients is not recorded, there is a unique choice of origin–the only choice for which ϕ is invariant under $P\bar{1}$. The situation concerning the handedness of the coordinate system is slightly different. Since ϕ is invariant under $P\bar{1}$ it must be that $\phi_n = \phi_{-n}$. Therefore, the two functions ϕ_n and ϕ_{-n} that result from the two choices of handedness are the same function and so the choice of handedness does not

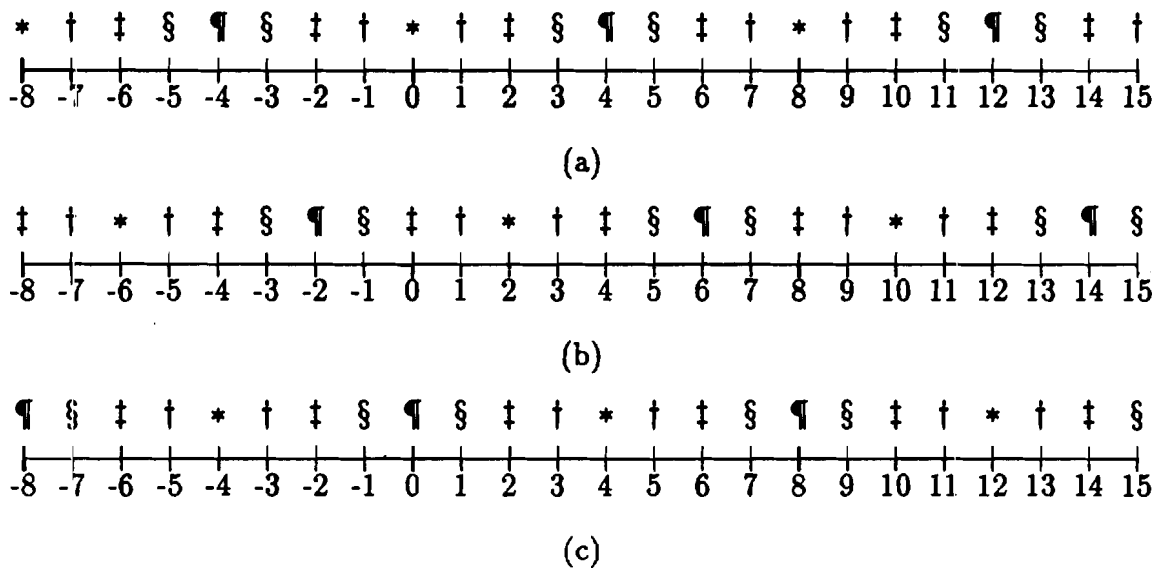


Figure 2: Symmetry breaking in Pi for L even, specifically $L = 8$: different symbols represent values that need not be equal. (a) Example of an invariant signal. (b) Example translation, specifically by 2, of the signal from (a) demonstrating the lack of invariance in the translated signal. (c) Example demonstrating that translation of the signal from (a) by half the period results in a different, but still invariant, signal.

matter. Therefore an exact estimator which knows that the signal is invariant under the $P\bar{1}$ space group with L odd does not need to use symmetry breaking. This is demonstrated in a numerical example in Section 3.

Now consider L even. In that case $L - 2$ of the points in the unit cell are paired in orbits of length 2 therefore sharing the same value and two points in the unit cell are isolated in separate orbits of length 1 therefore each having unique values. First consider the choice of origin location (Figure 2). If ϕ' is some translation of ϕ by an interval not equal to L or $L/2$ then ϕ' will typically not be invariant for the same reasons as for the case of L odd. The special case not considered for L odd is the case of translation by $L/2$. Let $\phi'_n = \phi_{n+L/2}$. Then $\phi'_n = \phi'_{n+L}$ is obvious and also $\phi'_{-n} = \phi_{-n+L/2} = \phi_{+n-L/2} = \phi_{+n+L/2} = \phi'_n$. Therefore ϕ'_n is a different, but still invariant, signal. The situation concerning the handedness of the

coordinate system is exactly the same as for L odd. Therefore an exact estimator which knows that the signal is invariant under the $P\bar{1}$ space group with L even continues to require at least a limited form of symmetry breaking in order to distinguish between ϕ_n and $\phi_{n+L/2}$. Because L even, relative to L odd, has this complication of requiring symmetry breaking even for an exact estimator, all of the calculations in this paper are for the L odd case.

In this problem the signals ϕ are 0-1. Therefore fortuitous equality of the values of ϕ at locations which are not in the same orbit is common. However, it seems unlikely that this will influence the need for symmetry breaking to a significant degree.

3 Introduction to the Numerical Example

The purpose of this section is to describe the numerical example that is studied with several algorithms in the course of this paper and to describe the results achieved using three basic estimators on this example. Two of the basic estimators are exactly-computed conditional mean estimators which differ only in the extent of their a *priori* knowledge concerning the symmetry. The first estimator, denoted "E-$P\bar{1}$", includes symmetry knowledge that the signal is invariant under $P\bar{1}$ (and therefore under $P1$ since $P1$ is a subgroup of Pi). The second estimator, denoted "E-$P1$", includes symmetry knowledge only that the signal is invariant under $P1$. E-P1 was used in Ref. [2, Section 6]. The third estimator, denoted 'A^n, is the approximate estimator with Problem 2 asymptotics from Ref [2, Section 6] which, like E-$P1$, includes symmetry knowledge only that the signal is invariant under $P1$.

The a *priori* Hamiltonian H^{apriori} used to generate and analyize the synthetic data in this example is the crystallographically motivated H^{aprion} presented in Ref. [2, Section 2]. Specifically, w_1 and w_2 in Eq. 1 have the form $w_1 = 0$ and $w_2(n_1, n_2) = \frac{1}{2}(\tilde{w}_2(n_1, n_2) + \tilde{w}_2(n_2, n_1))$ where $\tilde{w}_2(n_1, n_2)$ is defined by

$$\tilde{w}_2(n_1, n_2) = 0 \; n_1 \neq 0, \; \forall n_2$$

$$\tilde{w}_2(0, n_2) = \begin{cases} p_1, & 1 \le |n_2| < l_1 \\ \text{pa}, & l_1 \le |n_2| < l_2 \\ 0, & \text{otherwise} \end{cases}$$

and where $p_1 > 0$, $p_2 < 0$, and $1 \le l_1 \le l_2$. The range of atomic bond lengths that occur with high probability is $[l_1, l_2)$. This Hamiltonian can be used for any dimensional lattice. As discussed in Section 1, the formulae and examples in this paper are all one dimensional.

As discussed in Section 1, the formulae and examples in this paper are all for the space group $P\bar{1}$. This Hamiltonian assigns an energy to any lattice configuration. Nothing in this H^{apriori} guarantees that the lattice configuration will obey the Pi space group symmetry. Therefore, in order to compute realizations that obey this symmetry, a modified Metropolis algorithrn which incorporates the symmetry as a hard constraint is used. Specifically, rather than flipping the random variable ϕ_n at site n, all the random variables $\phi_{n_1}, \phi_{n_2}, \ldots$at sites $n_1, n_2, \ldots$that lie on a particular orbit are simultaneously flipped. Therefore, if the initial configuristion of the lattice is invariant under the $P\bar{1}$ space group symmetry, then all later configuristions are also invariant.

Comparison of the performance of the various approximate estimators with the exact estimators requires using a small lattice because the performance statistics are computed by the Monte Carlo method and the calculation of an individual estimate for the exact estimators is done by exhaustive enumeration. Therefore a one-dimensional lattice with period L = 17 is used. The remaining parameters in $H^{apriori}$ are $l_1 = 3$, $l_2 = 5$, $p_1 = 1.5$, and pa = –0.5.

The simulated data are produced in three steps. First N = 1000 configurations of the field ϕ are produced by the modified Metropolis algorithm with the parameters given above for H^{apriori}. In running the algorithm, the first 200000 configurations are discarded and then every 10000 th configuration is retained. Then the observational transformation (Fourier transform followed by the magnitude squared operation) is performed for each configuration. There are no parameters for this step. Finally, independent zero-mean Gaussian

pseudorandom variables with variance σ^2 are added to the Fourier coefficients for each configuration. The only parameter is σ^2 and for a particular data set σ^2 is constant for all Fourier coefficients. A range of σ^2 is considered–see the figures.

The fields ϕ that result from the simulation typically have four or five occupied sites. Therefore the Fourier coefficients typically do not exceed 4 or 5 in magnitude. Since real crystallographic data has 1 to 3 percent errors [11, p. 193], it is the performance of estimators at the low to moderate levels of a (i.e., $\sigma \leq .75$) that is most iimportant for the crystallographic application.

In general all of the numerical calculations discussed in this paper concern the performance of .matched estimators. That is, the parameters in the estimator match the parameters used to generate the synthetic data. In addition, typically the estimator has additional parameters that are described for each particular calculation.

Performance statistics are computed by Monte Carlo on N simulated data sets for a given choice of parameters such as a^2. Two measures of performance are considered. Both measures are expectations which are approximately computed by averaging the results of the N trials. Weighting by the probability mass function is not necessary since the configurations are drawn from the probability mass function.

Let $\hat{\phi}_n$ be an estimate of ϕ_n. Because the phases of the Fourier coefficients of 4 are not measured, $\hat{\phi}_{n+n_0}$ (translation by n_0) or $\hat{\phi}_{-n}$ (inversion through the origin) are equally satisfactory estimates. Therefore in this section min means a minimization over a possible inversion through the origin and a translation applied to $\hat{\phi}_n$.

The l_p norm is denoted $\|x\|_p = (\sum_n |x_n|^p)^{1/p}$. The first performance measure is the expected value of the l_2 norm of the difference between the true and reconstructed signals after a possible translation and reflection in order to achieve the best match, i.e., $E(l_2) = \mathrm{E} \min ||\phi - \hat{\phi}||_2$. In the estimators that guarantee to provide an estimate that satisfies the space group symmetry, the minimum for this minimization problem is often attained at $n_0 = 0$ because the true configuration is invariant under the symmetry and only the $n_0 = 0$

translation results in an estimate that is also invariant under the symmetry. It is, however, conceivable that on some occasions a shifted estimate which is no longer invariant will be chosen because it is a better match. This would typically happen at low signal to noise ratios when the estimator is performing poorly.

Let $\mu(\phi, \hat{\phi})$ be the minimum number of lattice sites where $\phi_n \neq \hat{\phi}_n$ and the minimum is taken over a possible inversion through the origin and a translation of $\hat{\phi}$. Note that $\min ||\phi - \hat{\phi}||_2^2 = \min ||\phi - \hat{\phi}||_1 = \mu(\phi, \hat{\phi})$. Therefore the performance results for mean squared error, mean absolute error, and mean number of lattice site differences are all the same.

The second measure, denoted f_{perfect}, is the probability of an error-free estimate, i.e., $\phi_n = \hat{\phi}_n$ for all n, again modulo inversion and translation. That is, $f_{\text{perfect}} = \Pr(\mu(\phi, \hat{\phi}) = 0) = E\delta_{\mu(\phi,\hat{\phi}),0}$ where throughout this paper $\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. The same comments regarding the minimization apply here also.

The first goal of the numerical work described in this paper is to demonstrate the increase in performance that is achieved by using E-$P\bar{1}$ rather than E-$P1$. That is, the first goal is to denionstrate the value of the additional symmetry information. Figures 3 and 4 each have three traces showing the performance of estimators E-$P\bar{1}$, E-$P1$, and A. All three estimators are matched to the synthetic data. Estimator E-$P\bar{1}$ has in addition q = 0 for the symmetry breaking parameter (i.e., no symmetry breaking whatsoever). Estimator E-$P1$ has in addition q = 1.0 for the symmetry breaking parameter and $\psi_n = n$ for the kernel of the symmetry breaking function. Estimator A has in addition q = 1.0 for the symmetry breaking parameter, $\psi_n^{\text{i.c.}} = \begin{cases} 0, & n = O \\ (L-n)/L, & n \neq 0 \end{cases}$ the initial condition for the symmetry breaking kernel (which makes $H^{\text{s.b.}}$ equal to the first moment of the field ϕ), $\gamma_1 = \gamma_2 = \gamma_3 = 1.0$ for the symmetry breaking optimization criteria, $\|\psi_n^{\text{i.c.}}\|_2$ as the target for the l_2 norm of ψ in the symmetry breaking optimization criteria, $\chi = 1.0$, $\lambda = 1.0$, and $\beta = 1.0$. For E-$P1$ and A these are the same parameters used in Ref. [2, Section 6].

Clearly, knowledge that the signal is invariant is valuable. For instance, in both performance measures, E-$P\bar{1}$ provides essentially perfect performance for a < 3 while for E-$P1$ the

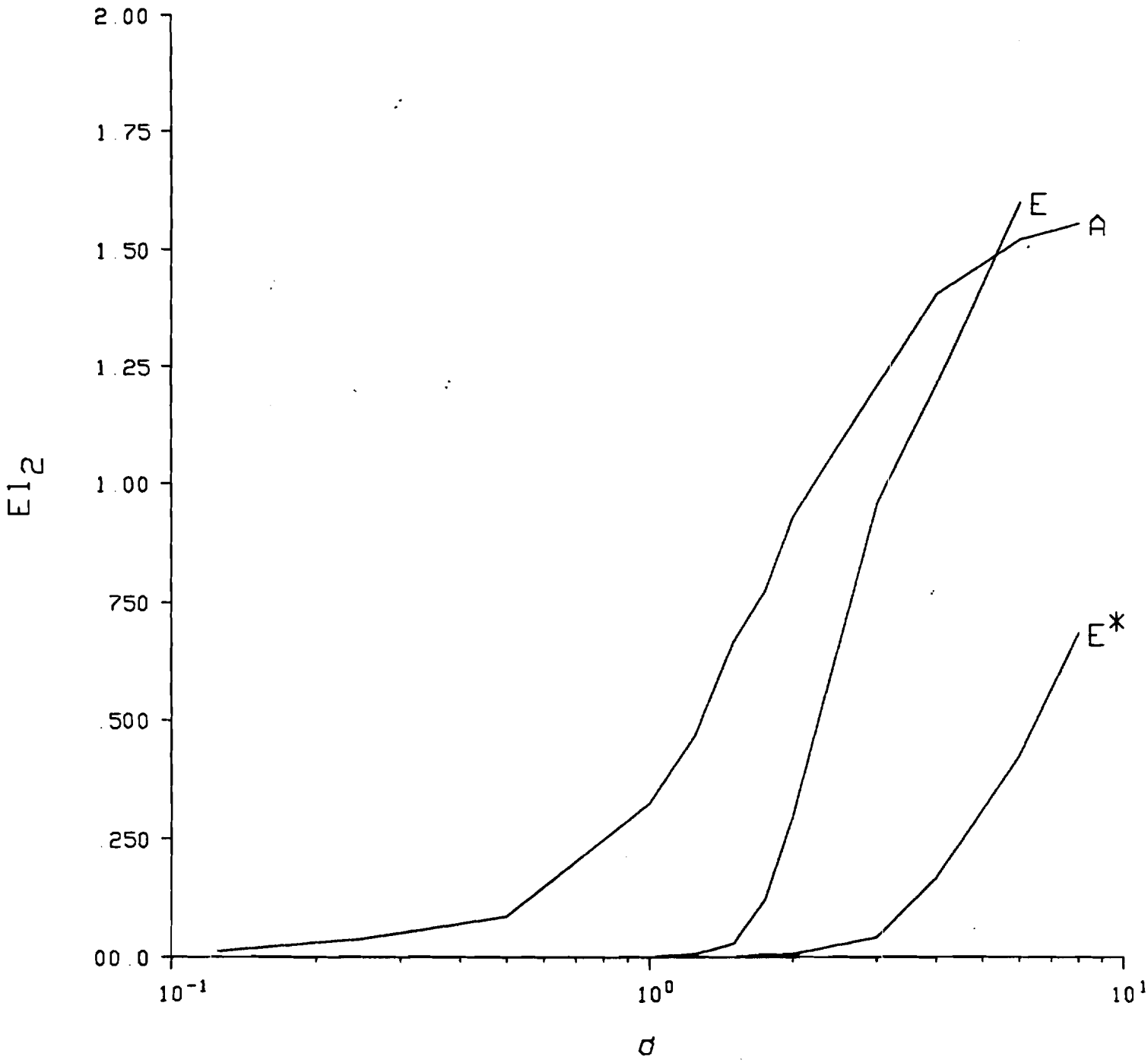


Figure 4: Estimator performance statistics: $E(l_2)$ versus σ for E-$P\bar{1}$ (trace "E^*"), E-P1 (trace "E"), and A (trace "A^n).

corresponding region is $a < 1.5$. Furthermore, for $a > 3$ the performance of $E\text{-}P\bar{1}$ declines at a slower rate than does the performance of $E\text{-}P1$ for $a > 1.5$.

The second goal of the'numerical part of this paper is to demonstrate an estimator that as near as possible closes the gap in Figures 3 and 4 between A, which has no knowledge of the invariance, and $E\text{-}P\bar{1}$, which has complete knowledge of the invisriance. It is not possible to close this gap completely. However, using knowledge of the invariance, it proves possible to develop a practical approximate estimator that, over a large range of σ^2, provides performance meeting or exceeding that provided by the impractical exact estimator $E\text{-}P1$ which lacks knowledge of the invariance.

In Section 1 it is claimed that symmetry breaking is not required if space group information is fully exploited. This fact is demonstrated in this example since $E\text{-}P\bar{1}$ achieves the indicated performance without any symmetry breaking.

The traces for $E\text{-}P1$ and A in Figures 3 and 4 correspond to Ref. [2, Figures 5 and 6] but are computed on different synthetic data sets which are not statistically equivalent to the synthetic data sets of Ref. [2, Figures 5 and 6] because the new data is guaranteed to satisfy the $P\bar{1}$ symmetry. Figures 5 and 6 compare the performances of $E\text{-}P1$ and A on the two data sets. For both estimators the performance on the guaranteed-symmetric data set is slightly superior. For A this superiority is maintained throughout the entire range of observation noise variance σ^2 while for E-P1 it is present only for low to moderate σ^2. An understanding of this characteristic of A might lead to changes that would improve its performance.

As described in Ref. [2, Section 4], the numerical optimization of the kernel of the symmetry breaking function for A was done using a multidimensional downhill simplex method [12, Section 10.4 pp. 305-309] applied to the ψ_n starting from an initial condition for which the symmetry breaking Hamiltonian was proportional to the first moment of the field ϕ. The same technique is used for all of the other approximate estimators described in this paper with the change that sometimes the multidimensional downhill simplex method is started

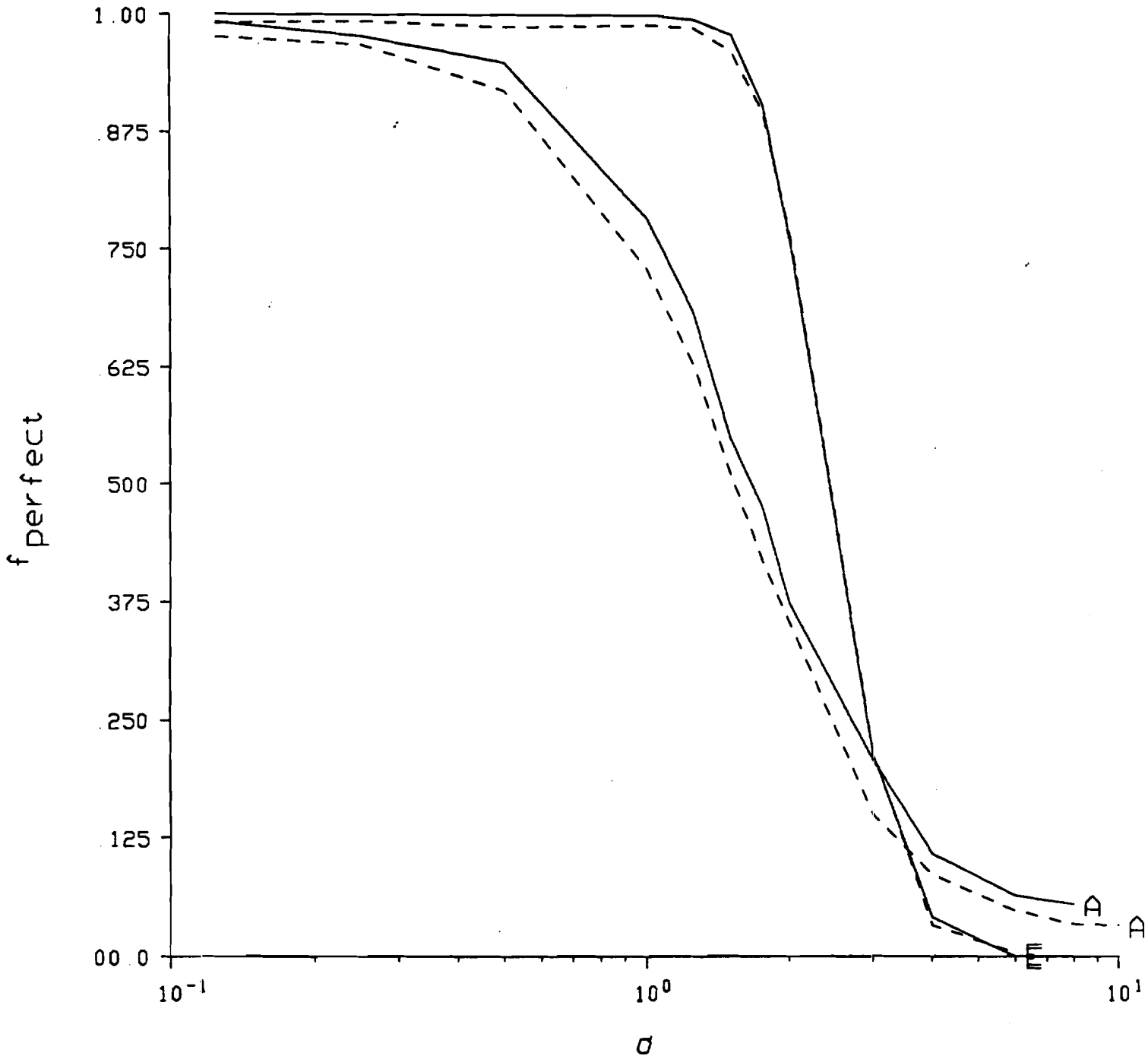


Figure 5: Estimator performance statistics: f_{perfect} versus a for E-$P1$ (trace "E") and A (trace "A"). Solid lines: data that is invariant under the actions of space group $P\bar{1}$. Dashed lines: data that is invariant only under the actions of space group $P1$.

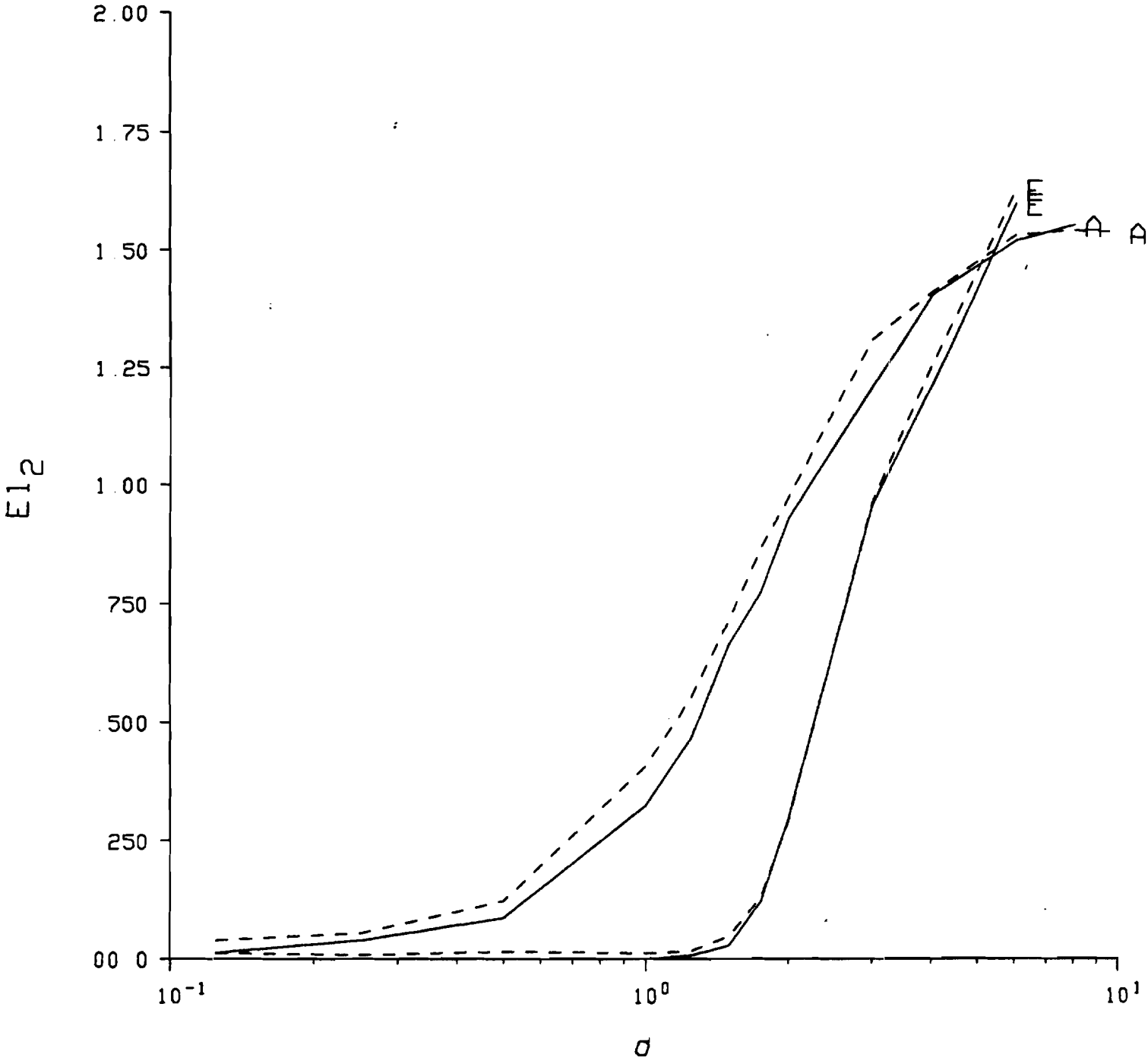


Figure 6: Estimator performance statistics: $E(l_2)$ versus σ for E-$P1$ (trace "E") and A (trace "A"). Solid lines: data that is invariant under the actions of space group $P\bar{1}$. Dashed lines: data that is invariant only under the actions of space group $P1$.

from a fixed number of randomly chosen initial conditions and the best of the results taken as the optimal ψ. If random initial conditions are used, they are always vectors whose components are independent identically distributed pseudorandom variables uniformly distributed over $[-1, 1)$. The number and type of initial conditions are described in later sections.

Recall that in the cost function for the optimal selection of ψ there is a penalty on deviation of the l_2 norm of ψ from a target value. The default for this target value is the same value used in Ref. [2, Section 4] which is the l_2 norm of $\psi_n^{i.c.} = \begin{cases} 0, & n = 0 \\ (L-n)/L, & n \neq 0 \end{cases}$.

4 Approach 1: space groups via averaging and soft constraints

Approach 1, where the space group symmetry is accounted for by averaging and soft constraints, is described in this section. The signal, in d-dimensions, is invariant under the actions of some space group denoted $\mathcal{G}$. The basic idea, as described in Section 1, is to replace $\mathcal{G}$ by the subgroup $P1$. The resulting signal reconstruction problem was solved [1, 2]. Then tht: information provided by knowledge that the signal is invariant under $\mathcal{G}$ is added back into the signal reconstruction algorithm using two methods which are denoted "averaging" and "soft constraints".

The averaging method is described first. Recall that A, mirroring the optimal estimator, operates in two steps: first compute m_n, an approximation to $E(\phi_n|y)$, and then compute the estimate $\hat{\phi}_n$ by thresholding m_n at $1/2$. If $\mathcal{G}$ is replaced by $P1$ then m_n and therefore $\hat{\phi}_n$ are typically not invariant under the actions of $\mathcal{G}$. The averaging method is to replace m_n by the average of m_n over the orbit of $\mathcal{G}$ than includes location n. This method fixes two problems. First the averaged m_n and therefore the reconstruction $\hat{\phi}_n$ that results from thresholding the averaged m_n are invariant under the actions of $\mathcal{G}$ as desired. Second, the signal to noise ratio is improved, though not to the degree possible if the information contained in the invariance under $\mathcal{G}$ is used from the start as a hard constraint.

The use of the averaging method is not trivial, however, because the $P1$ estimator is not guaranteed to give an estimate that has the same coordinate system as the true field. Specifically, the estimate could be translated and/or reflected through the origin relative to the true field. Therefore it would be foolish to average over the orbits in the true field coordinate system.

In light of the difficulties described in the previous paragraph, the averaging method is applied in two steps: first estimate the coordinate system in m_n and second average over the orbits in this new coordinate system. The only information concerning the coordinate system comes from the invariance of the true signal under the actions of $\mathcal{G}$. Therefore, in this paper choose as an estimate of the coordinate system that coordinate system which makes m_n and the orbit-averaged m_n most nearly equal in the l_2 sense.

Specifically, define $m_n^{n_0,s} = m_{sn+n_0}$ where $s \in \{\pm 1\}$. Therefore $m_n^{n_0,s}$ is a translated and reflected version of m_n. Let $\mathcal{O}_n$ be the orbit that includes location n and let $|\mathcal{O}_n|$ be the length of the orbit. Then the orbit-averaged $m_n^{n_0,s}$, denoted $\bar{m}_n^{n_0,s}$, is

$$\bar{m}_n^{n_0,s} = \frac{1}{|\mathcal{O}_n|} \sum_{n' \in \mathcal{O}_n} m_{n'}^{n_0,s},$$

the l_2 difference between $m_n^{n_0,s}$ and $\bar{m}_n^{n_0,s}$, denoted $C_4(n_0, s)$, is

$$C_4(n_0, s) = \sum_{n \in U} (m_n^{n_0,s} - \bar{m}_n^{n_0,s})^2,$$

and n_0 and s are chosen as the location of the minimum of $C_4(n_0, s)$:

$$n_0^*, s^* = \arg\min_{n_0,s} C_4(n_0, s).$$

The averaged estimate of $E(\phi_n|y)$, denoted m_n^a, is

$$m_n^a = \bar{m}_n^{n_0^*,s^*}.$$

Finally, the reconstruction, denoted $\hat{\phi}_n^a$, is

$$\hat{\phi}_n^a = \begin{cases} 1, & m_n^a \geq 1/2 \\ 0, & m_n^a < 1/2 \end{cases}.$$

For the case of one-dimensional signals invariant under the actions of $P\bar{1}$, it is necessary to consider translations n_0 but not reflections $s = -1$ because $P\bar{1}$ itself includes reflections. In addition, the criteria $C_4(n_0, s = 1)$ simplifies to

$$C_4(n_0, s = 1) = 1/2 \sum_{n=0}^{L-1} (m_n^{n_0, s=1} - m_{L-n}^{n_0, s=1})^2.$$

It is also necessary to consider when the averaging is done. One could average m, at every iteration of the optimization for the symmetry breaking kernel before the optimization criteria is computed and during the computation of the estimate using the optimized symmetry breaking kernel. Alternatively, one could average m_n only during the computation of the estimate using the optimized symmetry breaking kernel. The first approach might be expected to give better results than the second but runs the risk of making the kernel optimization difficult because it makes the optimization problem nondifferentiable. In fact, as seen in Section 5, given the numerical optimization tools available, the second approach provides superior performance.

The second method is the soft constraints method. Recall that the estimate computed by estimator A depends on an optimization criteria for the symmetry breaking kernel. In Ref [2, Section 3] a three term choice for the criteria was motivated. Since the optimization is numerical, it is relatively straightforward to add an additional term, though it is possible that such a term would make the optimization more difficult. The soft constraint method is to include the invariance of 4 under the actions of $\mathcal{G}$ as a soft constraint by modifying this criteria. Specifically, the modification is to add a term which penalizes deviations of ϕ from invariance.

The same general point of view as in the averaging method is used. Specifically, define the orbit-averaged m_n, denoted $\bar{m}_n$, as

$$\bar{m}_n = \frac{1}{|\mathcal{O}_n|} \sum_{n' \in \mathcal{O}_n} m_{n'},$$

the l_2 difference between m_n and $\bar{m}_n$, denoted C_4, as

$$C_4 = \sum_{n \in U} (m_n - \bar{m}_n)^2,$$

and add a term $\gamma_4 C_4$ to the cost function developed in Ref. [2, Section 3] where γ_4 is a constant weight. For the case of one dimensional signals invariant under the actions of $P\bar{1}$ this simplifies to the addition of the term

$$\gamma_4/2 \sum_{n=1}^{\frac{L-1}{2}} (m_n - m_{L-n})^2$$

to the cost function of Ref. [2, Section 3].

Note that while invariance failure is penalized, it is not forbidden. Thai; is, the constraint is soft rather than hard. Therefore estimates computed using the soft constraints method alone will typically not exhibit $P\bar{1}$ symmetry. For that reason, the soft constraints method is always used in combination with the averaging method.

5 Approach 1: numerical results

In this section the performances of three estimators based on Approach 1 are presented and compared with the three basic estimators. The problem and the three basic estimators (E-$P\bar{1}$, E-$P1$, and A) are discussed in Section 3.

The three estimators based on Approach 1 are

1. Averaging applied at the end only (denoted 'Al-end").
2. Averaging applied at the end and at every ψ optimization iteration (denoted "*A1-always*").
3. Cost function modification plus averaging applied at the end only (denoted "*A1*").

The primary results are shown in Figures 7 and 8. The three basic estimators are described

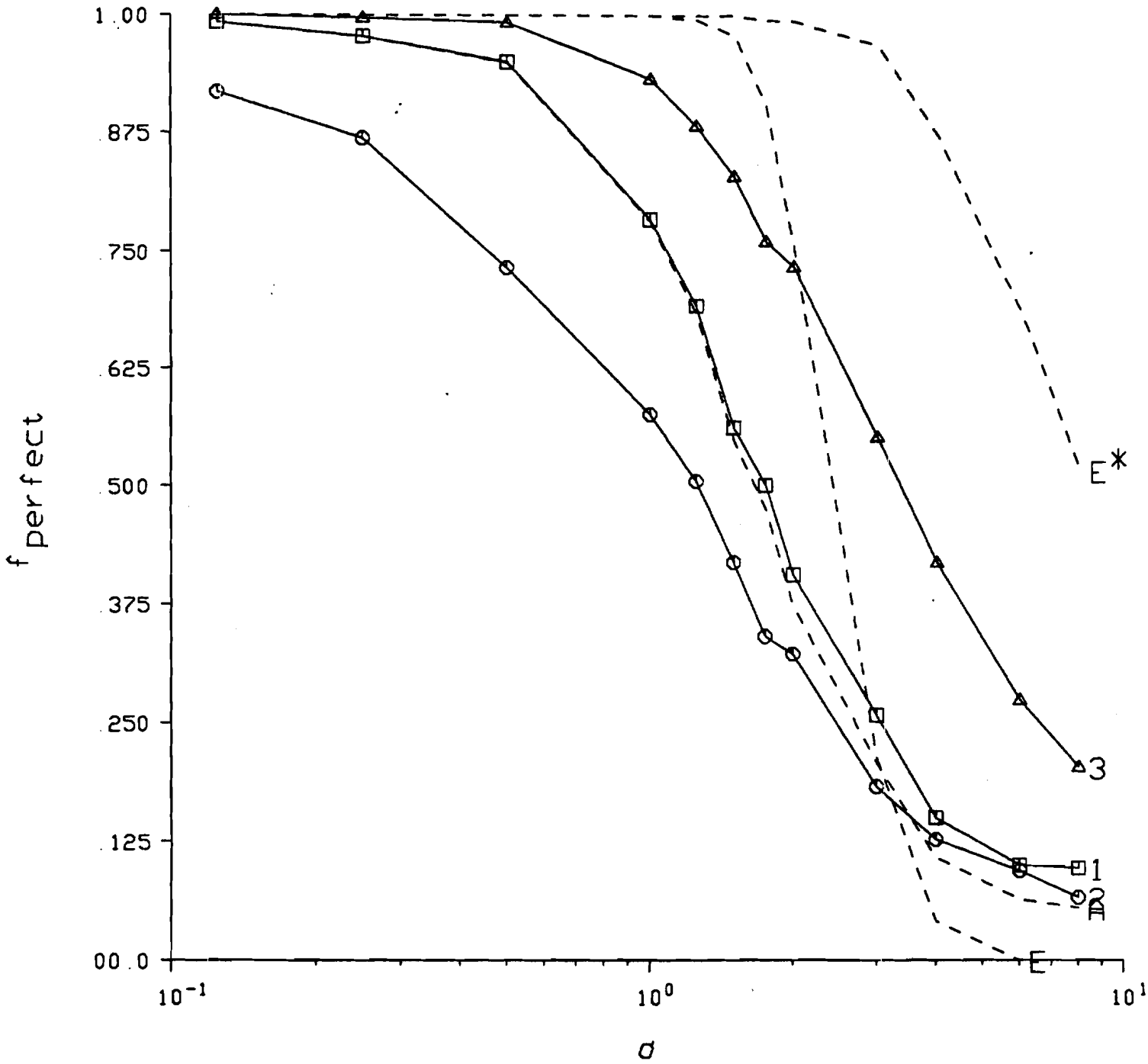


Figure 7: Estimator performance statistice: f_{perfect} versus σ for the three basic estimators of Section 3 (dotted lines) md Al-end (trace "1"), Al-always (trace "2"), and $A1$ (trace "3").

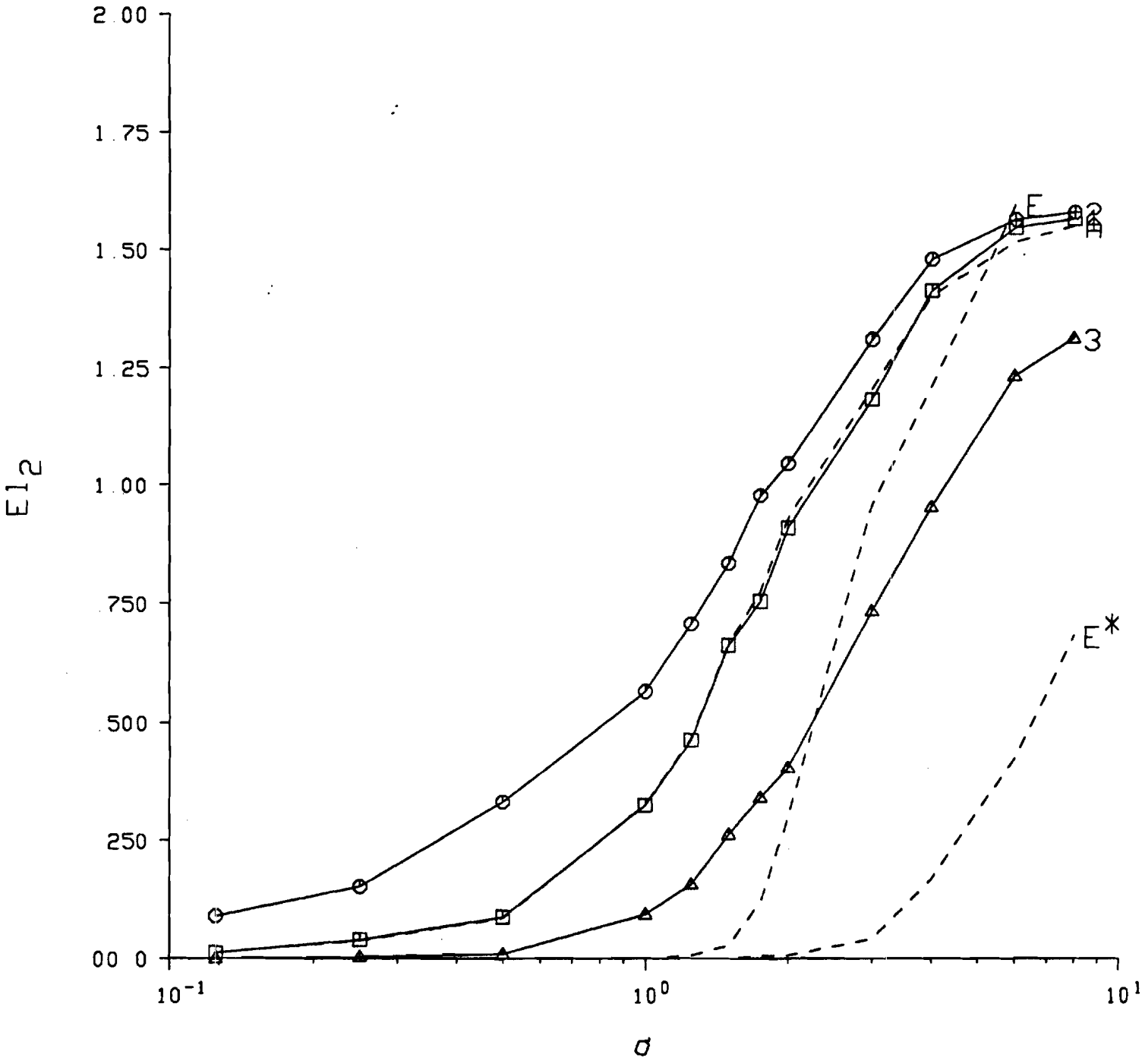


Figure 8: Estimator performance statistics: $E(l_2)$ versus a for the three basic estimators of Section 3 (dotted lines) and Al-end (trace "1"), *A1-always* (trace "2"), and *A*1 (trace "3").

in Section 3. The three new estimators are matched to the synthetic data. Estimators Al-end and Al-always have in addition q = 1.0 for the symmetry breaking parameter, $\epsilon = \begin{cases} 0, & n = O \\ (L-n)/L, & n \neq 0 \end{cases}$ as the initial condition for the 11, optimization, $\gamma_1 = \gamma_2 = \gamma_3 = 1.0$ and $\gamma_4 = 0$ for the ψ optimization criteria, the default target (see Section 3) for $\|\psi\|_2$ in the ψ optimization criteria, $\chi = 1.0$, $\lambda = 1.0$, and $\beta = 1.0$. Estimator $A1$ has in addition $q = 1.0$ for the symmetry breaking parameter, 50 independent choices of random initial conditions for the ψ optimization (see Section 3), $\gamma_1 = \gamma_2 = \gamma_3 = 1.0$ and $\gamma_4 = 4.0$ for the 11, optimization criteria, the default target (see Section 3) for $\|\psi\|_2$ in the ψ optimization criteria, $\chi = 1.0$, $\lambda = 1.0$, and $\beta = 1.0$.

Estimators Al-end and A provide essentially the same performance: Estimator Al-end outperforms A only slightly, only at low signal to noise ratios, and only in the f_{perfect} performance measure. The essentially equal performance of these two estimators tends to indicate that the errors made by A are rather global in nature rather than isolated errors at single lattice sites since averaging pairs of symmetry related conditional mean estimates has little effect on performance.

Estimator Al-always provides uniformly poor performance. This is probably due to the fact that the 11, optimization criteria in Al-always is noncontinuous, the optimization technique is of a down-hill search nature, and only one initial condition is considered. Rather than explore noncontinuous optimization techniques, this estimator was clropped in favor of Estimator Al.

Estimator $A1$ provides uniformly superior performance. At high signal to noise ratios it equals E-$P1$ (the exact estimator without knowledge that the signal is invariant under $P\bar{1}$) while at moderately low to low signal to noise ratios it substantially outperforms E-$P1$.

In summary, the averaging and soft constraints methods, as combined in Al, are able to extract a significant fraction of the performance increase available due to the knowledge that the field ϕ is invariant under the symmetry. The example used here is simple-one dimensional with $P\bar{1}$ symmetry. Therefore it is important to emphasize that these ideas extend without

significant modification to two and three dimensions and complicated symmetries. The only information needed to implement the estimator is knowledge of the orbits of the space group and this information is tabulated for all two and three dimensional space groups in, for example, Ref. [4].

6 Approaches 2 and 3: Introduction

In Approaches 2 and 3 the space group is accounted for by viewing it as a hard constraint on the ϕ_n variables. Due to the hard constraint, only a subset of the ϕ_n variables can be set inclependently. A valid subset is exactly a valid fundamental domain for the space group. In both Approaches 2 and 3 the estimates are computed for ϕ_n where n ranges over a fundamental domain. The remaining values of ϕ_n are then set by the constraint.

The estimation calculation continues to use the spherical model. The difference between Approaches 2 and 3 is whether the spherical model is applied before (Approach 2) or after (Approach 3) the constraints implied by the space group. Equivalently, the distinction is whether the spherical model is applied to the entire unit cell (Approach 2) or only to one fundamental domain (Approach 3).

Relative to the calculations of Ref. [1], which were reused in Approach 1, the calculations for Approaches 2 and 3 follow the same principles but are quite different in details and results. Specifically, for Approach 2 there is a large set of critical points each of which makes a contribution to the value of the integral. These contributions can be analytically summed. On the other hand, for Approach 3 there is only a single critical point but its location cannot be determined analytically except in a special case.

Because the principles of the calculations are the same as in Ref. [I], the calculation is divided into the same steps to the extent possible. 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[n] (see Section 8) in this

choice of variables. This is the reason for the care in choosing H^{apriori} and $H^{\text{s.b.}}$ as described in Ref. [1]. Second, two approximations are introduced to address two different problems. First, the zero-one nature of the MRF lattice variables is 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 $\sigma_k \downarrow 0$.

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 priorimodel $H^{apriori}$ is progressively forgotten. In the second approximation ("Problem 2"), $H^{apriori} \uparrow \infty$ 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 $H^{\text{s.b.}}$ is introduced and $H = H^{\text{apriori}} + H^{\text{obs}} + H^{\text{s.b.}}$ is 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). Step 3 in Ref. [1, Section 6], writing the Fourier coefficients in terms of magnitude and phase variables, and Step 4 in Ref. [1, Section 7], exact evaluation of the phase variable integrals, are greatly changed because the Fourier coefficients of a function invariant under $P\bar{1}$ turn out to be real. In Step 3 in the present paper the Hamiltonian is additively partitioned as before, notation is defined to make the correspondence with Ref. [1] as close as possible, and the spherical model constraints for Approaches 2 and 3 are compared. In Step 4 in the present paper the conditional mean integrals rue written out and

the need for symmetry breaking in Approach 2 is demonstrated even if the integrals could be evaluated exactly.

The conditional mean integrals over the red-valued Fourier coefficients are performed by the asymptotic approximation. The two different asymptotic approximations are defined (Step 5). In both cases the integral is of Laplace type and the integration region is a manifold due to the spherical model constraint. The two different asymptotic approximations turn out to diflkr 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). Through Step 7 the calculations for Approaches 2 and 3 are parallel and are presented jointly.

The critical point for the asymptotic small noise approximation is computed (Step 8). Formal calculations rather than rigorous proofs of the asymptotic formulae are provided. Several steps are required. An important part of the calculation which is common to Approaches 2 and 3 is the computation of a second order asymptotic expansion of a multivariable integral when there is a single critical point that is internal to the region of integration. This calculation is performed in Appendix 26. The plan for using the results in Appendix 26 is outlined (Step 9). The necessary Taylor expansion results and quantities derived from them are computed in Appendices 22, 23, 24, and 25 (Step 10). 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; ratios of the asymptotic expansions, which are the approximations to $E(\Phi_k|y)$, are computed; and the inverse Fourier series and nonlinear thresholding' leading to the estimate $\hat{\phi}_n$ of the field ϕ_n are described (Step 11). Following the sections deriving tht: estimator for Approach 2 (3) there is a section describing the results of numerical experiments using the estimator.

7 Spherical Model

The purpose of this section is to describe the two different spherical models used respectively in Approaches 2 and 3. Because Approach 2 (3) focuses on the unit cell (fundamental domain), the subscript "uc" ("fd") is used to label functions pertaining to Approach 2 (3).

For Approach 2 first the spherical model and then the symmetry constraint are applied. The starting point is the sums that express the estimator that is exact, though not aware of the symmetry. These sums are

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

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

where the sums are over configurations of the lattice in the entire unit cell, that is, n $\in$ $\{0,\ldots,L-1)$.

The summations of Eqs. 2 and 3 over configurations of the binary-valued ϕ_n for n $\in$ $\{0,\ldots,L-1)$ are written as integrals over R^L with a weighting function

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

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 $\vec{\phi}_{\text{uc}} = (\phi_0, \ldots, \phi_{L-1})$. The spherical model approximation is to replace $w_{\text{uc}}^{\text{exact}}$, which constrains $\vec{\phi}_{\text{uc}}$ to lie at the corners of an L-dimensional hypercube, by $w_{\text{uc}}^{\text{spherical}}$, which is defined to constrain $\vec{\phi}_{\text{uc}}$ to lie on the hypersphere circumscribed around the hypercube. Specifically, unchanged from Ref. [1], the spherical model is $w_{\text{uc}}^{\text{spherical}} = \delta(C_{\text{uc}}(\phi))$ where

$$C_{\text{uc}}(\phi) = \sum_{n=0}^{L-1} \phi_n(\phi_n - 1). \tag{4}$$

For Approach 2, the application of the symmetry constraint, which follows the application of the apherical model, occurs with the change to Fourier coordinates and is described in Section 8.

For Approach 3 first the symmetry constraint and then the spherical :model are applied. The starting point is Eqs. 2 and 3 which express the estimator that is exact, though not aware of the symmetry. These sums are over configurations of the lattice! in the entire unit cell. Application of the symmetry constraint reduces the sums to sums over configurations of the sublattice contained in the fundamental domain. The natural fundamental domain for Pi, as discussed in Section 2, is $\{0, \ldots, \frac{L-1}{2}\}$. The new equations are

$$Z^{\text{exact}}(y) = \sum_{\phi} e^{-H_{\text{fd}}(y,\phi)} \tag{5}$$

$$E^{\text{exact}}(\phi_n|y) = \frac{1}{Z^{\text{exact}}(y)} \sum_{\phi} \phi_n e^{-H_{\text{fd}}(y,\phi)} \tag{6}$$

where the sums are only over configurations of the sublattice $n \in \{0, \ldots, \frac{L-1}{2}\}$ and H has been changed to H_{fd} to indicate that it is now a function of a limited set of 4.

The summations of Eqs. 5 and 6 over configurations of the binary-valued ϕ_n for $n \in \{0, \ldots, \frac{L-1}{2}\}$ are written as integrals over $R^{\frac{L-1}{2}+1}$ with a weighting function

$$w^{\text{exact}} = \prod_{n=0}^{\frac{L-1}{2}} \delta(\phi_n(\phi_n - 1)).$$

Let $\vec{\phi}_{\text{fd}} := (\phi_0, \ldots, \phi_{\frac{L-1}{2}})$. The spherical model approximation is to replace $w_{\text{fd}}^{\text{exact}}$, which constrains $\vec{\phi}_{\text{fd}}$ to lie at the corners of an $\frac{L-1}{2}+1$-dimensional hypercube, by $w_{\text{fd}}^{\text{spherical}}$, which is defined to constrain $\vec{\phi}_{\text{fd}}$ to lie on the hypersphere circumscribed around the hypercube. Specifically, the spherical model is $w_{\text{fd}}^{\text{spherical}} = \delta(C_{\text{fd}}(\phi))$ where

$$C_{\text{fd}}(\phi) = \sum_{n=0}^{\frac{L-1}{2}} \phi_n(\phi_n - 1). \tag{7}$$

Note how the number of sites included in C_{fd} and therefore w_{fd} is roughly 1/2 the number in C_{uc} and therefore w_{uc} which implies that the w_{fd} approximation is more accurate.

In future sections, any material not specifically labeled Approach 2 versus Approach 3 applies to both and in particular the notation $w^{\text{spherical}}$ and C applies to either Approach 2 with $w^{\text{spherical,uc}}$ and C_{uc} or Approach 3 with $w^{\text{spherical,fd}}$ and C_{fd}. This completes Step 1.

8 Fourier coordinates

In this section 'the Hamiltonian and spherical model constraint are expressed in terms of the Fourier coefficients of ϕ, denoted Φ, rather than ϕ. This is the natural set of coordinates because the Hamiltonian, which is quartic in terms of either ϕ or 9, is diagonal in terms of 9. That is, in terms of Φ, the Hamiltonian does not have any cross product terms, e.g., terms such as $\Phi_{k_1}\Phi_{k_2}$ with $k_1 \neq k_2$.

The first task is to determine how $\phi_n \in \mathrm{R}$ and ϕ invariant under $P\bar{1}$ constrains the Fourier coefficients Φ of ϕ. First recall a standard fact:

Fact 1 *$\phi_n \in R$ if and only if $\Phi_k = \Phi^*_{L-k}$.*

The desired result is a generalization of this standard result to the case where ϕ is invariant under $P\bar{1}$. The generalization is:

Fact 2 *$\phi_n \in R$ and ϕ invariant under* Pi *if and only if $\Phi_k = \Phi_{L-k}$ and $\Phi_k \in R$.*

The demonstration of this fact, a straightforward calculation, is omitted.

In comparison with Step 2 in Ref. [1, Section 6], the present calculation is changed since Φ now has fewer independent degrees of freedom. Specifically, since 9 is real and conjugate symmetric it is convenient to take $\Phi_0 = 34$, $\Phi_1 = \Re\Phi_1$, ..., $\Phi_{\frac{L}{2}} = \Re\Phi_{\frac{L}{2}}$ as the $\frac{L}{2}+1$ independent degrees of freedom for L even and $\Phi_0 = \Re\Phi_0$, $\Phi_1 = \Re\Phi_1$, ..., $\Phi_{\frac{L-1}{2}} = \Re\Phi_{\frac{L-1}{2}}$ as the $\frac{L-1}{2}+1$ independent degrees of freedom for L odd. As explained in Section 2, all of the calculations in this paper are for the L odd case.

The statement that these are all of the possible degrees of freedom in Φ carries with it the information that ϕ is invariant under the actions of the $P\bar{1}$ space group. For the estimator

of Approach 3 this is not any additional information since that constraint has already been imposed in terms of ϕ. However, for Approach 2, this constraint has not been previously applied. Therefore, as described in Section 7, the symmetry constraint is applied in the process of transforming from ϕ to Φ. Once this information has been applied, the Hamiltonians for Approachs 1 and 2 are again identical and the symbol "H" is used. To elaborate on this point (for L odd), for Approach 3 the original function $H(\phi_0,\ldots,\phi_{L-1})$ is transformed to $H'(\phi_0,\ldots,\phi_{\frac{L-1}{2}})$ by applying the symmetry constraint and then is transformed to $H''(\Re\Phi_0,\ldots,\Re\Phi_{\frac{L-1}{2}})$ by changing variables while for Approach 2 the same original function is transformed to $H^{\dagger}(\Re\Phi_0,\Re\Phi_1,\Im\Phi_1,\ldots,\Re\Phi_{\frac{L-1}{2}},\Im\Phi_{\frac{L-1}{2}})$ by changing variables and then is transformed to $H^{\dagger\dagger}(\Re\Phi_0,\ldots,\Re\Phi_{\frac{L-1}{2}})$ by applying the symmetry constraint. The result is that $H'' = H^{\dagger\dagger}$.

Define $\Phi_{r,k} = \Re\Phi_k$, $\Phi_{i,k} = \Im\Phi_k = 0$, $K_L = \{0,1,\ldots,\frac{L-1}{2}\}$, and $K_L^+ = \{1,\ldots,\frac{L-1}{2}\}$. Writing out the total Hamiltonian, using $\Phi_{r,k}$ rather than Φ_k in order to emphasize that Φ is real gives

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

For Approach 2, Eq. 4 implies

$$
C_{\mathrm{uc}} \;=\; \frac{1}{L}\Phi_{r,0}^2 - \Phi_{r,0} + \frac{2}{L}\sum_{k=1}^{\frac{L-1}{2}}\Phi_{r,k}^2.
$$

For Approach 3, the Fourier analysis and synthesis equations for a function ϕ_n that is invariant under the actions of $P\bar{1}$ can be written

$$\begin{aligned}
\Phi_k &= \phi_0 + \sum_{n=1}^{\frac{L-1}{2}} \phi_n 2\cos nk\frac{2\pi}{L} \\
&= \sum_{n=0}^{\frac{L-1}{2}} \phi_n(2-\delta_{n,0})\cos nk\frac{2\pi}{L} \\
\phi_n &= \frac{1}{L}\left(\Phi_0 + \sum_{k=1}^{\frac{L-1}{2}} \Phi_k 2\cos nk\frac{2\pi}{L}\right) \\
&= \frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0})\cos nk\frac{2\pi}{L}.
\end{aligned}$$

Using these formulae in Eq. 7 gives

$$\begin{aligned}
C_{\mathrm{fd}}(\phi) &= \sum_{n=0}^{\frac{L-1}{2}} \phi_n(\phi_n - 1) \\
&= \sum_{n=0}^{\frac{L-1}{2}} \phi_n^2 - \sum_{n=0}^{\frac{L-1}{2}} \phi_n \\
&= \sum_{n=0}^{\frac{L-1}{2}} \left(\frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0})\cos nk\frac{2\pi}{L}\right)\phi_n - \frac{1}{2}(\Phi_0+\phi_0) \\
&= \frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0}) \sum_{n=0}^{\frac{L-1}{2}} \phi_n \cos nk\frac{2\pi}{L} - \frac{1}{2}(\Phi_0+\phi_0) \\
&= \frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0})\frac{1}{2}(\Phi_k+\phi_0) - \frac{1}{2}(\Phi_0+\phi_0) \\
&= \frac{1}{2L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k^2(2-\delta_{k,0}) + \frac{1}{2L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0})\phi_0 - \frac{1}{2}(\Phi_0+\phi_0) \\
&= \frac{1}{2L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k^2(2-\delta_{k,0}) + \frac{1}{2L^2}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0}) \sum_{k'=0}^{\frac{L-1}{2}} \Phi_{k'}(2-\delta_{k',0}) \\
&\quad - \frac{1}{2}\Phi_0 - \frac{1}{2L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k(2-\delta_{k,0}) \\
&= \frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k^2\frac{2-\delta_{k,0}}{2} + 2\left(\frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k\frac{2-\delta_{k,0}}{2}\right)^2 - \frac{1}{2}\Phi_0 - \frac{1}{L}\sum_{k=0}^{\frac{L-1}{2}} \Phi_k\frac{2-\delta_{k,0}}{2} \\
&= \sum_{k=0}^{\frac{L-1}{2}} \Phi_k^2 v_k + 2\left(\sum_{k=0}^{\frac{L-1}{2}} \Phi_k v_k\right)^2 - \frac{1}{2}\Phi_0 - \sum_{k=0}^{\frac{L-1}{2}} \Phi_k v_k
\end{aligned}$$

$$= \Phi^T \operatorname{diag}(v_0, v_1, \ldots, v_{\frac{L-1}{2}})\Phi + 2\Phi^T v v^T \Phi - \frac{1}{2}\Phi_0 - v^T \Phi$$

where $v_k = (2 - \delta_{k,0})/(2L)$, k $\in K_L$; $v = (v_0, v_1, \ldots, v_{\frac{L-1}{2}})^T = (\frac{1}{2L}, \frac{1}{L}, \ldots, \frac{1}{L})^T$; and $\Phi = (\Phi_0, \Phi_1, \ldots, \Phi_{\frac{L-1}{2}})^T$. This completes Step 2.

Introduce a parameter β, analogous to inverse temperature in statistical mechanics, that allows the entire Harniltonian to be simultaneously scaled. Take advantage of the fact that the contribution to H of each Φ_k is additive by defining

$$-\beta h_k = a_{k,0} + a_{k,1}\Phi_{r,k} + a_{k,2}\Phi_{r,k}^2 + a_{k,4}\Phi_{r,k}^4$$

for any k $\in K_L$ where the $a_{k,j}$ definitions are stated in Appendix 20. (The definitions of both βh_k and $a_{k,j}$ are changed relative to Ref. [1, Section 6 and Appendix A], though the only change in the $a_{k,j}$ is in $a_{k,1}$ for k $\neq$ 0). Then,

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

Because of the new form for $-\beta h_k$ relative to Ref. [1, Section 6], there is no need to introduce rotated variables Φ'_k k $\in K_L^+$ and change to magnitude (r_k) and phase (θ_k) variables. However, in order to make the current equations as similar to the equations of Ref. [1] as possible, introduce the notation

$$r_k = \Phi_{r,k} = \Re\{\Phi_k\} = \Phi_k$$

emphasizing that r_k takes values in R not $R^+ \cup \{0\}$. In these variables the equations have the form

$$\begin{aligned}
-\beta h_k &= a_{k,0} + a_{k,1} r_k + a_{k,2} r_k^2 + a_{k,4} r_k^4 \\
C_{\text{uc}} &= \frac{1}{L} r_0^2 - r_0 + \frac{2}{L} \sum_{k=1}^{\frac{L-1}{2}} r_k^2 \\
C_{\text{fd}} &= \sum_{k=0}^{\frac{L-1}{2}} r_k^2 v_k + 2\left(\sum_{k=0}^{\frac{L-1}{2}} r_k v_k\right)^2 - \frac{1}{2} r_0 - \sum_{k=0}^{\frac{L-1}{2}} r_k v_k \\
&= r^T \operatorname{diag}(v_0, v_1, \ldots, v_{\frac{L-1}{2}}) r + 2 r^T v v^T r - \frac{1}{2} r_0 - v^T r.
\end{aligned}$$

It is helpful to better understand the difference between two constraints C_{uc} and C_{fd}. Both are quadratic forms. In the natural $r_k = \Phi_k$ coordinates, C_{uc} is diagonal while C_{fd} is not. In this and the following paragraphs these quadratic forms are transformed to standard form and their eigenvectors and eigenvalues are computed.

The transformation of C_{uc} to standard form is simple since the quadratic form is already diagonal. All that is required is to complete the square in r_0 with the result that

$$\begin{aligned} C_{uc} &= \frac{1}{L}\left(r_0 - \frac{L}{2}\right)^2 + \frac{2}{L}\sum_{k=1}^{\frac{L-1}{2}} r_k^2 - \frac{L}{4} \\ &= (r + d_{uc})^T \operatorname{diag}(\frac{1}{L}, \frac{2}{L}, \ldots, \frac{2}{L})(r + d_{uc}) - \frac{L}{4} \end{aligned} \tag{8}$$

where $d_{uc} = (-\frac{L}{2}, 0, \ldots, 0)^T$. Therefore there are $\frac{L-1}{2}$ eigenvalues with value $\frac{2}{L}$ and with an eigenvector subspace spanned by $e1, \ldots, e_{\frac{L-1}{2}}$, and a single eigenvalue with value $\frac{1}{L}$ and with eigenvector e_0 where $e_i \in R^{\frac{L-1}{2}+1}$ are the standard basis vectors numbered from 0 to $\frac{L-1}{2}$. That is, e_i is a vector of zeros except for a single 1 in component i where the first component is numbered 0 rather than the more conventional 1.

The transformation of C_{fd} to standard form requires the matrix version of completing the square. This formula, for any symmetric invertible matrix Σ, is $r^T Cr + b^T r = (r + d)^T \Sigma (r + d) - d^T \Sigma d$ where $d = \frac{1}{2}\Sigma^{-1} b$. For C_{fd} the matrix Σ has the form $\Sigma_{fd} = \operatorname{diag}(v_i) + 2vv^T$. Application of the Woodbury formula [12, p. 76] to compute Σ_{fd}^{-1} gives the result that $\Sigma_{fd}^{-1} = \operatorname{diag}(\frac{1}{v_i}) - (1, \ldots, 1)^T(1, \ldots 1$. Since $b_{fd} = (-\frac{1}{2}, 0, \ldots, 0)^T - v$, it follows that $d_{fd} = (-\frac{L}{2}, 0, \ldots, 0)^T$ and $d_{fd}^T \Sigma_{fd} d_{fd} = \frac{L+1}{8}$. Note that $d_{fd} = d_{uc}$. Combining these results gives

$$C_{fd} = (r + d_{fd})^T \Sigma_{fd} (r + d_{fd}) - \frac{L+1}{8}. \tag{9}$$

In order to compute the eigenvectors and eigenvalues of Σ_{fd}, begin by noting that the matrix

$$\Sigma_{fd} - \frac{1}{L} I = \operatorname{diag}(-\frac{1}{2L}, 0, \ldots, 0) + 2vv^T$$

is only of rank 2. Therefore, the first $\frac{L-1}{2} - 1$ eigenvalues are equal with common value denoted $\chi = \frac{1}{L}$. Furthermore the null space of this matrix, and hence the eigenvector subspace corresponding ta χ, is all vectors $\xi = (w_0, w_1, \ldots, w_{\frac{L-1}{2}})$ of the form $w_0 = 0$ and $w_1, \ldots, w_{\frac{L-1}{2}}$ such that $\sum_{k=1}^{\frac{L-1}{2}} w_k = 0$. One basis for the eigenvector subspace is

$$w^i = \sum_{j=1}^{\frac{L-1}{2}} e_j - \frac{L-1}{2} e_i$$

for $i \in \{1, 2, \ldots, \frac{L-1}{2} - 1\}$. A second basis is

$$w^i = e_j - e_i$$

for any fixed $j \in \{1, 2, \ldots, \frac{L-1}{2}\}$ and for $i \in \{1, 2, \ldots, j-1, j+1, \ldots, \frac{L-1}{2}\}$.

The remaining two eigenvalues and eigenvectors are more difficult to compute. From numerical experiments using MATLAB™[13] it appeared that the final two eigenvectors were of the form

$$\xi = (\alpha, \beta, \ldots, \beta)^T.$$

The corresponding eigenvalue χ is, by definition, the number such that ξ lies in the null space of $A - \chi I$, i.e., $(A - \chi I)\xi = 0$. Since

$$(A - \chi I)\xi = \begin{pmatrix} \left(\frac{1}{2L} - \chi\right)\alpha \\ \left(\frac{1}{L} - \chi\right)\beta \\ \vdots \\ \left(\frac{1}{L} - \chi\right)\beta \end{pmatrix} + \frac{2}{L^2} \begin{pmatrix} \frac{1}{2} \\ 1 \\ \vdots \\ 1 \end{pmatrix} \left(\frac{1}{2}\alpha + \frac{L-1}{2}\beta\right),$$

it is necessary and sufficient to require that α and β satisfy a homogeneous two-dimensional linear system which is

$$\begin{pmatrix} \frac{1}{2L} + \frac{1}{2L^2} - \chi & \frac{L-1}{2L^2} \\ \frac{1}{L^2} & \frac{1}{L} + \frac{L-1}{L^2} - \chi \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0.$$

The result for this 2 x 2 eigenvalue/eigenvector problem for the matrix

$$\begin{pmatrix} \frac{1}{2L} + \frac{1}{2L^2} & \frac{L-1}{2L^2} \\ \frac{1}{L^2} & \frac{1}{L} + \frac{L-1}{L^2} \end{pmatrix}$$

is that

$$\chi_{\pm} = \frac{1}{4L^2}\left[5L - 1 \pm \sqrt{(L-1)(9L-1)}\right]$$
$$\begin{pmatrix}\alpha_{\pm}\\ \beta_{\pm}\end{pmatrix} = \begin{pmatrix}1\\ \frac{3}{2} \pm \frac{1}{2}\sqrt{\frac{9L-1}{L-1}}\end{pmatrix}.$$

Returning to the original eigenvector/eigenvalue problem, the result in the 2×2 problem implies that the last two eigenvalues and eigenvectors of Σ_{fd} are

$$\chi_{\pm} = \frac{1}{4L^2}\left[5L - 1 \pm \sqrt{(L-1)(9L-1)}\right]$$
$$\xi_{\pm} = \left(1, \frac{3}{2} \pm \frac{1}{2}\sqrt{\frac{9L-1}{L-1}}, \ldots, \frac{3}{2} \pm \frac{1}{2}\sqrt{\frac{9L-1}{L-1}}\right)^T.$$

The $C_{uc} = 0$ constraint (Eq. 8) can be rewritten

$$\left(\frac{r_0}{\sqrt{2}} - \frac{L}{2\sqrt{2}}\right)^2 + \sum_{k=1}^{\frac{L-1}{2}} r_k^2 = \left(\frac{L}{2\sqrt{2}}\right)^2$$

which is a sphere of radius $\frac{L}{2\sqrt{2}}$ which is stretched and displaced frorn the origin in the $(1, 0, \ldots, 0)^T$ direction. Let T_{fd} be the matrix whose columns are an orthonormal set of eigenvectors of Σ_{fd}. The $C_{fd} = 0$ constraint (Eq. 9) can be rewritten

$$\left(T_{fd}^H(r + d_{fd})\right)^H \mathrm{diag}(1, \ldots, 1, L\chi_+, L\chi_-)\left(T_{fd}^H(r + d_{fd})\right) = \left(\frac{\sqrt{L(L-1)}}{2\sqrt{2}}\right)^2.$$

In the correct rotated coordinate system defined by the unitary matrix T_{fd}, this is a sphere of radius $\frac{\sqrt{L(L-1)}}{2\sqrt{2}}$ which is stretched in two coordinates and displaced from the origin. Thus these two constraints are quite similar. However, the fact that the natural coordinates for the C_{fd} constraint are not the coordinates in which the Hamiltonian is diagonal makes it much more difficult to solve for the critical point location analytically.. (The eigenvector-eigenvalue structure of C_{fd} is also used in Section 17 to chose initial conditions for a numerical computation). This completes Step 3.

9 Bayesian integrals

The central quantities in the Bayesian estimators described in this paper are the conditional means $E(\phi_n|y)$ and approximations to them. In this section approximations to the conditional means are expressed as multidimensional integrals using the spherical model integration. measure, the inability to compute these integrals in terms of standard functions is noted which motivates the asymptotic evaluation of these integrals, and for Approach 2 the role of symmetry breaking in the asymptotic evaluation is elucidated.

Invariance of ϕ under the actions of the $P\bar{1}$ space group guarantees that Φ is real and therefore, in comparison with Ref. [1, Section 7], there are no longer any angular integrals. The expressions for Approach 2 and Approach 3 are identical. Writing out the approximation under the spherical model to the partition function $Z^{\text{exact}}(y)$ (Eq. 5) gives

$$Z(y) = \int_{-\infty}^{+\infty} dr_0 \int_{-\infty}^{+\infty} dr_1 \cdots \int_{-\infty}^{+\infty} dr_{\frac{L-1}{2}} w^{\text{spherical}} e^{-\beta H} \tag{10}$$

Following Ref. [1], the mean of the field is computed in terms of the rnean of its Fourier coefficients. That is, an approximation under the spherical model to $E^{\text{exact}}(\Phi_k|y)$ rather than to $E^{\text{exact}}(\phi_n|y)$ is computed. For the mean of Φ_k, the integrand for Z is multiplied by

$$\Phi_k = r_k$$

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

$$E(\Phi_{r,k}|y) = \frac{1}{Z(y)} \int_{-\infty}^{+\infty} dr_0 \int_{-\infty}^{+\infty} dr_1 \cdots \int_{-\infty}^{+\infty} dr_{\frac{L-1}{2}} r_k w^{\text{spherical}} e^{-\beta H} \tag{11}$$

$$E(\Phi_{i,k}|y) = 0 \tag{12}$$

where k $\in K_L$. The remaining $E(\Phi_k|y)$ are specified by $\Phi_k = \Phi_{L-k}$, that is, $E(\Phi_k|y) = E(\Phi_{L-k}|y)$.

These integrals do not appear to be solvable in terms of standard functions. Therefore, as detailed in Sections 10, 13, and 16, an asymptotic evaluation is performed.

A major difference between Approaches 2 and 3 is the necessity of symmetry breaking in Approach 2 since without symmetry breaking the conditional expectations for $k \neq 0$ are identically 0. This fact is not an undesirable side effect of the asymptotic method of evaluation but is true in the original integral (Eq. 11). The remainder of this section describes the situation.

The absence of symmetry breaking corresponds to $\psi_n = 0$ for all $a \in \{0, \ldots, L-1\}$ and/or to $q = 0$. In this case, since $a_{k,1} = 0$ for all k except $k = 0$, $E(\Phi_{r,k}|y)$ for $k \# 0$ can be written as

$$\begin{aligned} E(\Phi_{r,k}|y) &= \frac{1}{Z(y)} \int_{-\infty}^{+\infty} dr_0 \int_{-\infty}^{+\infty} dr_1 \cdots \int_{-\infty}^{+\infty} dr_{\frac{L-1}{2}} r_k w_{uc}^{\text{spherical}} e^{-\beta H} \\ &= \frac{1}{Z(y)} \int_{-\infty}^{+\infty} dr_0 \int_{-\infty}^{+\infty} dr_1 \cdots \int_{-\infty}^{+\infty} dr_{\frac{L-1}{2}} r_k \delta\left(\frac{1}{L} r_0^2 - r_0 + \frac{2}{L} \sum_{l=1}^{\frac{L-1}{2}} r_l^2\right) \times \\ &\quad \times \exp\left(a_{0,1} r_0 + \sum_{l=0}^{\frac{L-1}{2}} \left(-a_{l,0} + a_{l,2} r_l^2 + a_{l,4} r_l^4\right)\right). \end{aligned}$$

Order the integrations so that the r_k integration is performed last and perform the other integrations in order to get

$$E(\Phi_{r,k}|y) = \int_{-\infty}^{+\infty} r_k f(r_k) dr_k \tag{13}$$

where $f(\cdot)$ is an even function because r_k only enters the delta and exponential functions through r_k^2 and the region of integration for r_l for $l \neq k$ does not involve r_k. Since f is even it follows that $r_k f(r_k)$ is odd. Since the region of integration is even it then follows that the integral is zero as claimed. This completes Step 4.

10 Asymptotics

The asymptotic ideas of Ref. [1, Section 8] are used to evaluate Eqs. 10 and 11. 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, it is

assumed that $\sigma_k^2 = \frac{1}{\lambda}\bar{\sigma}_k^2$ and $A \uparrow \infty$. The second limit, denoted Problem 2, combines the small observation noise limit with a proportional scaling of the a *priori* Hamiltonian. That is, it is assumed 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. This completes Step 5.

With the correct notation, the asymptotic evaluation of Eqs. 10 and 11 requires the asymptotic expansion of integrals of the form $\int_D a(x)e^{\lambda\gamma(x)}dx$ in the limit $\lambda \to \infty$ where γ is real, and D is all of $R^{\frac{L-1}{2}+1}$. This is a problem of Laplace type [14, Section 6.4 pp. 261–276]. Not only the order in A but also the numerical coefficient of the first nonzero term in the $A \to \infty$ asymptotic series is required.

The points where the exponent γ attains a global maximum, called critical points, play an important role in the large-A asymptotics because as $A \to \infty$ the entire contribution to the integral comes from a neighborhood of these points. Though it does not contribute to the determination of the critical points, the behavior of a (the nonexponential part of the integrand), especially the points at which a and perhaps its derivatives vanish, is also important because these points may, and in fact do, occur at the critical points. Therefore the following sections define notation so that the integrals are of this form (Section 11), locate the points where a vanishes (Section 11), and locate the critical points (Sections 12 and 15).

11 Asymptotics–notation

The first goal of this section is to define notation so that the partition function (Eq. 10) and conditionirl means (Eq. 11) can be written

$$\begin{aligned} Z(\lambda) &= \int g_Z w^{\text{spherical}} e^{-\lambda\beta H_\lambda} \\ E(\Phi_{r,k}|y)(\lambda) &= \frac{1}{Z(\lambda)} \int g_k w_{\text{spherical}} e^{-\lambda\beta H_\lambda} \quad k \in K_L. \end{aligned}$$

First define some quantities related to the exponent. Having introduced A and χ, 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 asymptotics), 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 20. The only differences relative to the definitions of Ref. [1, Appendix A] are in $b_{k,1c}$ for both Problem 1 and 2. The only difference between Problems 1 and 2 is the definition of these constants $b_{k,ns}$ and for both Problem 1 and Problem 2 it follows from the definitions that, as in Ref. [1],

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

Make! explicit the λ dependence of the exponent by defining

$$\begin{aligned} -\beta h_{k,0} &= b_{k,1c} r_k + b_{k,2c} r_k^2 \qquad (14) \\ -\beta h_{k,1} &= \begin{cases} -b_{0,0a} + b_{0,1b} r_0 + (b_{0,2a} + b_{0,2b}) r_0^2 - b_{0,4a} r_0^4, & k = 0 \\ -b_{k,0a} + (b_{k,2a} + b_{k,2b}) r_k^2 - b_{k,4a} r_k^4, & k \in K_L^+ \end{cases} \end{aligned}$$

so that

$$-\beta h_k = -\beta h_{k,0} - \lambda \beta h_{k,1}$$

and there is no other λ dependence in h_k. Define

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

(In comparison with Ref. [1, Section 9] these definitions are unchanged except that the irrelevant "r" subscript is removed since there are no longer any angular "0" variables and $b_{k,1c}$ for k $\in K_L^+$ are no longer hidden within Θ_k^0 and Θ_k^1 [1, Section 7 Eqs. 5 and 9]).

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

$$g_Z(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = \exp(\sum_{k=0}^{\frac{L-1}{2}} -\beta h_{k,0}(r_k)) \tag{15}$$

$$\begin{aligned} g_k(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) &= r_k \exp(\sum_{j=0}^{\frac{L-1}{2}} -\beta h_{k,0}(r_j)) \quad k \in K_L \\ &= r_k g_Z(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) \end{aligned} \tag{16}$$

which are all independent of A.

The second goal is to fix some notation concerning the critical point. This notation is carried over unchanged from Ref. [1, Section 9]. 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 always denotes a variable in $R^{\frac{L-1}{2}+1}$ while the variable $\bar{r}$ always denotes a variable in $R^{\frac{L-1}{2}}$. Components of the critical point ρ that are zero play an important role. Define

$$A_\rho = \{k \in K_L | \rho_k = 0\} \tag{17}$$

$$\bar{A}_\rho = \{k \in K_L^+ | \rho_k = 0\}. \tag{18}$$

Recall

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

Therefore, the integrals of Eqs. 10 and 11 are over the manifold defined by $C(r_0, r_1, \ldots, r_{\frac{L-1}{2}}) = 0$. (Compare with Ref. [1, Section 9] where the integration is only over a subset of the manifold). The implicit function theorem assures the existence in a neighborhood of ρ of a continuously differentiable function $\eta_\rho : R^{\frac{L-1}{2}} \to \mathrm{R}$ such that

$$C(\eta_\rho(\bar{r}), \bar{r}) = 0 \tag{19}$$

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

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

$$(F_\rho(\bar{r}))_k = \begin{cases} \eta_\rho(\bar{r}), & k = 0 \\ r_k, & k \in K_L^+ \end{cases} \tag{20}$$

Note that $F_\rho(\bar{\rho}) = \rho$. Note also that there are actually two η_ρ functions ($\eta_{\rho,\mathrm{uc}}$ and $\eta_{\rho,\mathrm{fd}}$) and therefore! two F_ρ functions ($F_{\rho,\mathrm{uc}}$ and $F_{\rho,\mathrm{fd}}$) corresponding to the two C functions (C_{uc} and C_{fd}). This completes Step 6.

The third goal is to state properties of the zeros of g and the derivatives of g. First, g_Z never vanishes. Second, g_k vanishes if and only if $r_k = 0$. This completes;Step 7.

Gaussian integrals play an important role. Define as in Ref. [1, Section 9]

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

which is the normalization factor for a Gaussian density with covariance matrix Q^{-1} (i.e., $p_{m,Q^{-1}}(r) = N(Q)\exp(-\frac{1}{2}(r-m)^T Q(r-\mathrm{m})))$. In addition, because it appears frequently throughout Approaches 2 and 3, define

$$\begin{aligned} f_k(r_k) &= \partial_{r_k}\partial_{r_k}\beta H_\lambda \\ &= -2(b_{k,2a} + b_{k,2b}) + 12 b_{k,4a} r_k^2. \end{aligned} \tag{21}$$

Finally, the invariance properties of $g_Z(\rho_0, \rho_1, \ldots, \rho_{\frac{L-1}{2}})$ under sign reversals on components ρ_k for k > 0 are important. Define $\rho'_0 = \rho_0$ and $\rho'_k = |\rho_k|$ for k > 0. Define

$$g_Z^{\mathrm{inv}}(\rho'_0) = \exp(-\beta h_{0,0}(\rho'_0)). \tag{22}$$

Then

$$\begin{aligned} g_Z(\rho_0, \rho_1, \ldots, \rho_{\frac{L-1}{2}}) &= \exp(\sum_{k=0}^{\frac{L-1}{2}} -\beta h_{k,0}(\rho_k)) \\ &= \exp(-\beta h_{0,0}(\rho_0))\exp(\sum_{k\in \bar{A}_\rho} -\beta h_{k,0}(\rho_k))\exp(\sum_{k\in K_L^+ - \bar{A}_\rho} -\beta h_{k,0}(\rho_k)) \\ &= \exp(-\beta h_{0,0}(\rho'_0))\exp(\sum_{k\in \bar{A}_\rho} -\beta h_{k,0}(0))\exp(\sum_{k\in K_L^+ - \bar{A}_\rho} -\beta h_{k,0}(\rho_k)) \\ &= g_Z^{\mathrm{inv}}(\rho'_0)\exp(\sum_{k\in K_L^+ - \bar{A}_\rho} -\beta h_{k,0}(\rho_k)) \end{aligned} \tag{23}$$

where $h_{k,0}(0) = 0$ and $\rho_0 = \rho'_0$ have been used.

12 Asymptotics–critical point for Approach 2

The critical point is the minimum of βH_λ. The definition of βH_λ is unchanged from Ref. [1, Section 9], including the same definitions for those $b_{k,ns}$ that enter βH_λ. Two minimization problems were discussed in Ref. [1, Section 10], specifically,

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

In Ref. [1], the solution of Opt 1 was required, but it was possible to show that any solution of Opt 2 reflected into the orthant $\{r_0 \in R\} \times \{r_k \geq 0 | k \in K_L^+\}$ was a solution of Opt 1 and that there were no solutions of Opt 1 that were not also solutions of Opt 2 (i.e., there were no solutions due to the boundary). In the present paper the solution of Opt 2 itself is required. Therefore, Ref. [1, Section 10] actually contains the needed results. However, in Ref. [1, Section 10] the reflection into $\{r_0 \in \text{R}) \text{ x } \{r_k \geq 0 | k \in K_L^+\}$ is performed in the process of computing the solution rather than computing the solution to Opt 2 and then reflecting. The change amounts to the introduction of a plus/minus sign in Ref. [1, Eqs. 23 and 24]. The new equations are

$$\rho_k(\tau) = \begin{cases} \pm\sqrt{\frac{(b_{k,2a}+b_{k,2b})-\tau\frac{2}{L}}{2b_{k,4a}}}, & \tau < \frac{L}{2}(b_{k,2a}+b_{k,2b}) \\ 0, & \tau \geq \frac{L}{2}(b_{k,2a}+b_{k,2b}) \end{cases} k \in K_L^+ \cap B \tag{24}$$

$$\tau \geq \frac{L}{2} b_{k,2b} \; \forall k \in K_L^+ - B$$

$$\rho_k(\tau) = \begin{cases} 0, & \tau > \frac{L}{2} b_{k,2b} \\ \text{arbitrary}, & \tau = \frac{L}{2} b_{k,2b} \end{cases} k \in K_L^+ - B \tag{25}$$

where B := {k $\in K_L^+$ | an observation was taken at frequency k or frequency L − k). Ref. [1, Eqs. 19 and 21], which are,

$$-b_{0,1b} - 2(b_{0,2a}+b_{0,2b})\rho_0 + 4b_{0,4a}\rho_0^3 + \tau(\frac{2}{L}\rho_0 - 1) = 0 \tag{26}$$

$$\frac{1}{L}\rho_0^2 - \rho_0 + \frac{2}{L}\sum_{k=1}^{\frac{L}{2}-1}\rho_k^2 = 0, \tag{27}$$

complete the solution. Eq. *25* amounts to $\rho_k(\tau) = 0$ k $\in K_L^+ - B$ since the probability that $\tau = \frac{L}{2}b_{k,2l}$ is zero.

Note how Eq. *26* is independent of $\rho_k(\tau)$ for k $\in K_L^+$ and how Eq. *27* depends on $\rho_k(\tau)$ for k $\in K_L^+$ only through $\rho_k(\tau)^2$ and is therefore independent of the plus/minus sign in Eq. 24 or the arbitrary/nonnegative distinction in Eq. *25.* Therefore, *dl* solutions can be generated with the following steps: first require $\rho_k(\tau) \geq 0$ for all k $\in K_L^+$, second solve for ρ_0 and τ, and third generate all the other solutions by flipping the signs of ρ_k for k $\in K_L^+$ while keeping ρ_0 and τ fixed. Denote the solution from the combination of the first and second steps by p'.

Recall the definition of $\bar{A}_\rho$ from Eq. 18, specifically, $\bar{A}_\rho = \{k \in K_L^+ | \rho_k = 0\}$. In terms of $\bar{A}_\rho$, the number of solutions that are generated is $2^{|K_L^+ - \bar{A}_\rho|}$. Each of these solutions is a critical point. Because βH_λ depends on r_k k $\in K_L^+$ only through r_k^2 it follows that

$$\beta H_\lambda(\rho) = \beta H_\lambda(\rho'). \tag{28}$$

That is, the exponent has the same value at each of the critical points and therefore none of the critical points are dominant over others and therefore all need to be included in the solution. This completes Step 8 for Approach *2.*

13 Asymptotics–formulae for Approach 2

In this section, asymptotic formulae for Approach *2* are presented. As in Ref. [1, Section 11], the calculations are formal, in the spirit of Ref. [14], rather than rigorous proofs. In all cases the lowest order nonzero term in the asymptotic expansion is computed. In d l but the final case, the lowest order nonzero term is of order 0 while in the find c u e it is of order *2.*

Formulae for second order asymptotic expansions of multivariable integrals in the case where there is a single critical point which is internal to the region of integration are given in Appendix 26. An important difference between the present calculation and the calculation

of Ref. [1] is that the region of integration for the present problem has no finite boundaries. Formulae for the necessary derivatives and derived quantities are given in Appendices 22, 23, 24, and 25.

Recall from Section 12 that there are $2^{|K_L^\dagger - A_\rho|}$ critical points for this problem. Because the critical points are all isolated, following Ref. [14], the expansion of the entire integral is the sum of the contributions due to each critical point. Furthermore, the contribution of a particular critical point p can be computed by restricting the region of integration to a neighborhood of p which excludes all other critical points and then applying the single-critical point formulae given in Appendix 26. Therefore, the plan has four steps:

1. Decompose the integral into a sum of critical point contributions.
2. Perform the r_0 integration for each individual critical point contribution using the 6-function of the spherical model.
3. Approximate the remaining integrations for each individual critical point contribution using the asymptotic formulae from Appendix 26.
4. Sum the individual critical point contributions.

This completes Step 9 for Approach 2.

The normalizer Z of the probability density (equivalently, partition function) is

$$Z = \int g_Z(r)\delta(C_{\text{uc}}(r))e^{-\lambda\beta H_\lambda(r)}dr.$$

Apply the four step plan from the previous paragraph. Decompose the integral into a sum over the critical points p which are all related to ρ' as described in Section 12 to get

$$Z = \sum_\rho Z_\rho \tag{29}$$

where

$$Z_\rho = \int_{r_0\in(\rho_0-\epsilon,\rho_0+\epsilon)} dr_0 \prod_{j=1}^{\frac{L-1}{2}} \int_{r_j\in(\rho_j-\epsilon,\rho_j+\epsilon)} dr_j g_Z(r)\delta(C_{\text{uc}}(r))e^{-\lambda\beta H_\lambda(r)}$$

and ϵ describes the neighborhood which is taken sufficiently small such that it contains only the single critical point p. The value of an individual contribution Z_ρ can be computed by first performing the r_0 integration (taking advantage of the δ-function of the spherical model constraint:) and then using the formulae of Appendix 26. Because g_Z never vanishes, the leading nonzero term of the asymptotic expansion is the zeroth order term (i.e., the term proportional to "q" in Eq. 84). Specifically, in terms of L, defined in Appendix 22 Eq. 49,

$$\begin{aligned} Z_\rho &= \prod_{j=1}^{\frac{L-1}{2}} \int_{r_j \in (\rho_j - \epsilon, \rho_j + \epsilon)} dr_j g_Z(F_\rho(\bar{r})) e^{-\lambda\beta H_\lambda(F_\rho(\bar{r}))} \\ &\approx e^{-\lambda\beta H_\lambda(\rho)} N^{-1}(\lambda L_\rho) g_Z(\rho) \end{aligned}$$

where the multiplicative factor derived from the $\delta(C_{uc}(r))$ integration which is common to all integrals and therefore cancels from the ratios has been dropped. More specifically, this factor derives from the fact that under suitable limitations on g one has $\int f(x)\delta(g(x))dx = \frac{f(g^{-1}(0))}{g'(g^{-1}(0))}$ where g' is the derivative of g. The denominator is common to all integrals, cancels from the ratios of interest, and therefore can be dropped. This completer the first three steps in the computation of Z.

The fourth and final step in the computation of Z is to sum Z_ρ over the critical points. As shown in Eq. 23, the term $g_Z(\rho)$ is the product of two terms, one term (g_Z^{inv} defined in Eq. 22) which is invariant with respect to the sign changes in $k \in K_L^+$ that generate p from ρ' and one term which is not invariant. Specifically,

$$g_Z(\rho) = g_Z^{inv}(\rho_0') \exp\Big(\sum_{k \in K_L^+ - \bar{A}_\rho} -\beta h_{k,0}(\rho_k)\Big).$$

Furthermore, from Appendix 21, $\det L_\rho$ is invariant under there sign changes and therefore

$$N(\lambda L_\rho) = N(\lambda L_{\rho'}).$$

Finally, as noted in Eq. 28,

$$-\beta H_\lambda(\rho) = -\beta H_\lambda(\rho').$$

Therefore:,

$$Z_\rho \approx e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0') \exp\Big(\sum_{k\in K_L^+ - \tilde{A}_{\rho'}} -\beta h_{k,0}(\rho_k)\Big). \tag{30}$$

Summing Eq. 30 over all of the critical points (Eq. 29) gives

$$\begin{aligned} Z &\approx e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0') \sum_\rho \exp\Big(\sum_{k\in K_L^+ - \tilde{A}_{\rho'}} -\beta h_{k,0}(\rho_k)\Big) \qquad (31)\\ &= e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0') \prod_{k\in K_L^+ - \tilde{A}_{\rho'}} \sum_{\rho_k\in\{\pm\rho_k'\}} \exp(-\beta h_{k,0}(\rho_k)) \\ &= e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0') \prod_{k\in K_L^+ - \tilde{A}_{\rho'}} \exp(b_{k,2c}\rho_k'^2) 2\cosh(b_{k,1c}\rho_k') \qquad (32) \end{aligned}$$

where the definition (Eq. 14) of $-\beta h_{k,0}$ has been used.

Next compute $E(\Phi_k|y)$ when $k \in K_L^+ - \tilde{A}_{\rho'}$. (Note that $k = 0$ is not in this set). Therefore, $\rho_k \neq 0$ which implies that g_k is not zero at the critical point. Therefore the contribution of interest is again the zeroth order contribution.

The plan for the Z calculation can be followed unchanged to Eq. 31 which now takes the form

$$\begin{aligned} ZE(\Phi_k|y) &\approx e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0') \sum_\rho \exp\Big(\sum_{l\in K_L^+ - \tilde{A}_{\rho'}} -\beta h_{l,0}(\rho_l)\Big)\rho_k \\ &= e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0')\rho_k' \left(e^{-\beta h_{k,0}(\rho_k')} - e^{-\beta h_{k,0}(-\rho_k')}\right) \times \\ &\quad \times \prod_{l\in K_L^+ - \tilde{A}_{\rho'} - \{k\}} \sum_{\rho_l\in\{\pm\rho_l'\}} \exp(-\beta h_{l,0}(\rho_l)) \\ &= e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) g_Z^{\text{inv}}(\rho_0')\rho_k' \exp(b_{k,2c}\rho_k'^2) 2\sinh(b_{k,1c}\rho_k') \times \\ &\quad \times \prod_{l\in K_L^+ - \tilde{A}_{\rho'} - \{k\}} \exp(b_{l,2c}\rho_l'^2) 2\cosh(b_{l,1c}\rho_l'). \end{aligned}$$

Therefore:,

$$E(\Phi_k|y) \approx \frac{\rho_k' \sinh(b_{k,1c}\rho_k')}{\cosh(b_{k,1c}\rho_k')}.$$

Next compute $E(\Phi_0|y)$, i.e., the $k = 0$ case. Assume $\rho_0 \neq 0$. Then g_0 is nonzero everywhere so that the contribution of interest is again the zeroth order contribution.

The plan for the Z calculation can be followed to Eq. **32** which takes the form

$$ZE(\Phi_0|y) \approx e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) \rho'_0 g_Z^{\text{inv}}(\rho'_0) \times \tag{33}$$
$$\times \prod_{k \in K_L^+ - A_{\rho'}} \exp(b_{k,2c}\rho_k'^2) 2\cosh(b_{k,1c}\rho'_k)$$

which **implies** that

$$E(\Phi_0|y) \approx \rho'_0. \tag{34}$$

Finally compute $E(\Phi_k|y)$ when k $\in A_{\rho'}$. Therefore $\rho_k = 0$ which implies that $g_k = r_k g_Z$ is zero at the critical point. As in Ref. [1], the case $0 \in A_{\rho'}$ is not considered because $0 \in A_{\rho'}$ implies $\rho_k = 0$ for all k $\in K_L$.

In the previous three calculations, I performed the r_0 integration and then found that a zeroth order asymptotic expansion was nonzero. The reason is that **the nonexponential** part of the integrand (i.e., $g_Z(F_\rho(\bar{r}))$ or $g_k(F_\rho(\bar{r}))$) did not vanish at the critical point. For the present case, after performing the r_0 integration, a higher order asymptotic expansion is required. The first order expansion, as described in Appendix 26, is always **zero** by symmetry for this **particular** type of integral. However, the second order expansion is in general nonzero.

The **first** two steps (decomposition into individual critical point contributions and performing the r_0 integrations) are unchanged. The third step involves the calculation of the second order terms in **Eq.** 84. The terms $J_{1,a}$ and $J_{1,b}$ are zero because they are proportional to $g_k(\rho)$ ("q^n in **Eq.** 84) which is zero. Furthermore, as calculated in Appendix 24 **(Eq. 73)**, $J_{1,c} = 0$. However, as calculated in Appendix 25 (Eq. **78**), $J_{1,d}$ is nonzero, specifically,

$$J_{1,d} = g_Z(\rho) \frac{b_{k,1c}}{f_k(0) + \tau_L^4}.$$

Using Eq. 78 in Eq. 84 and then **Eq. 23** gives the result

$$ZE_\rho(\Phi_k|y) \approx e^{-\lambda\beta H_\lambda(\rho)} N^{-1}(\lambda L_\rho) \frac{1}{\lambda} g_Z(\rho) \frac{b_{k,1c}}{f_k(0) + \tau_L^4}$$
$$= e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) \frac{1}{\lambda} g_Z^{\text{inv}}(\rho'_0) \times$$
$$\times \exp\Big(\sum_{l \in K_L^+ - A_\rho} -\beta h_{l,0}(\rho_l) \Big) \frac{b_{k,1c}}{f_k(0) + \tau_L^4}.$$

This result can be summed over p in a fashion analogous to the summation of Z_ρ over p to yield

$$ZE(\Phi_k|y) = e^{-\lambda\beta H_\lambda(\rho')} N^{-1}(\lambda L_{\rho'}) \frac{1}{\lambda} g_Z^{\text{inv}}(\rho_0') \frac{b_{k,1c}}{f_k(0) + \tau \frac{4}{L}} \times$$
$$\times \prod_{l \in K_L^+ - \bar{A}_{\rho'}} \exp(b_{l,2c}\rho_l'^2) 2\cosh(b_{l,1c}\rho_l').$$

Finally, dividing through by the approximation to Z gives

$$E(\Phi_k|y) \approx \frac{1}{\lambda} \frac{b_{k,1c}}{f_k(0) + \tau \frac{4}{L}}$$

where f_k is defined in Eq. 21.

In this and the preceding sections one method is described for performing the calculations needed in Approach 2. A different method is to reduce the region of integration for the original integral (Eq. 11) to $r_k \in R^+ \cup \{0\}$ for k $\neq$ 0. This is similar to performing the sum over the critical points. Then, with this reduced region of integration,,the critical point calculations of Ref. [1, Section 10] can be used unchanged. However, in this second method, the critical point could fall on the boundary of the region of integration and this complication outweighs the advantages of this method.

The final two steps in the estimator are

1. to compute $m_n \approx E(\phi_n|y)$ from $M_k \approx E(\Phi_k|y)$ by computing the inverse Fourier series of M_k and
2. to compute the estimate $\hat{\phi}_n$ of the field ϕ_n by thresholding m_n at 1/2: $\hat{\phi}_n = \begin{cases} 1, & m_n \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$

Phase estimates, if desired, can be computed by computing the phase of the Fourier coefficients of the estimate $\hat{\phi}_n$.

In preparation for the numerical results, note that $b_{k,ns}$ (defined in Appendix 20) depend on Ψ only through $\Re\Psi$ (recall that ψ is real so that Ψ_0 is guaranteed to be real). Therefore, without loss of generality, it is possible to assume that $\Im\Psi_k = 0$ for all k. Since ψ_n is real, it is already guaranteed that $\Psi_k = \Psi_{L-k}^*$. Therefore, with this assumption, it follows that

$\Psi_k = \Psi_{L-k}$ and $\Psi_k \in$ R. Application of Fact 2 (Section 8) leads to the conclusion that ψ is invariant under $P\bar{1}$. Therefore, for L odd, an independent parameterization of ψ is $\psi_0, \ldots, \psi_{\frac{L-1}{2}}$. This completes Step 11 for Approach 2.

14 Numerical results–Approach 2

In this section the performance of the estimator based on Approach 2, denoted A2, is presented and compared with four alternative estimators. The problem and three of the alternative estimators–the basic estimators E-$P\bar{1}$, E-$P1$, and A–are discussecl in Section 3. The fourth alternative estimator, A1 based on Approach 1, is discussed in Sections 4 and 5.

The primary results are shown in Figures 9 and 10. Parameters for the three basic estimators are described in Section 3. Parameters for $A1$ are described in Section 5. Estimator A2 is matched to the synthetic data. It has in addition q = 1.0 for the symmetry breaking parameter, 50 independent choices of random initial conditions for the ψ optimization (see Section 3), $\gamma_1 = \gamma_2 = \gamma_3 = 1.0$ for the ψ optimization criteria, the default target (see Section 3) for $\|\psi\|_2$ in the ψ optimization criteria, $\chi = 0.5$, $\lambda = 1.0$, and $\beta = 1.0$.

In terms of the f_{perfect} performance measure, $A1$ and A2 provide similar performance. At low to moderate observation noise variance (σ^2) A2 has a slight advantage while at high σ^2 $A1$ has a modest advantage. The same is true in terms of the El_2 performance measure though at low to moderate σ^2 the performance of A2 at best equals that of $A1$ rather than exceeding it slightly. Recall (Section 3) that it is the low to moderate levels of σ^2 that are relevant to the crystallography application.

However, there is a second aspect that consistently favors A2. Specifically, as discussed in Section 13, for A2 the symmetry breaking kernel ψ can be parameterized by $\frac{L-1}{2} + 1$ real numbers. On the other hand, the kernel ψ for $A1$ requires a full L real numbers. Therefore the symmetry breaking optimization for A2 occurs in a space of essentially half the dimension of the optimization for A1. It is anticipated that the dimension of the parameterization of ψ

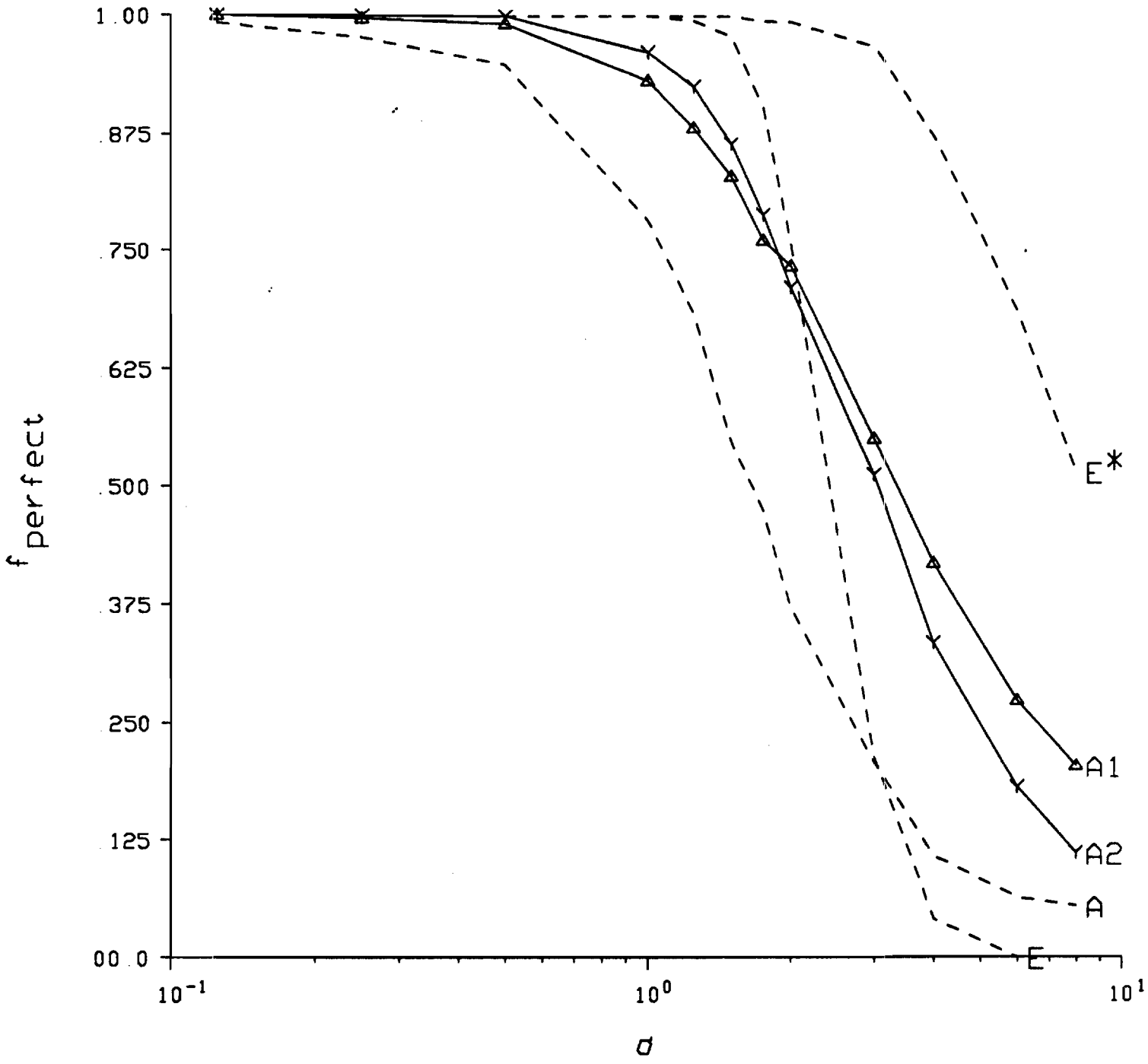


Figure 9: :Estimator performance statistics: $f_{perfect}$ versus σ for the three basic estimators of Section 3 (dotted lines), $A1$ (solid line labeled "$A1$"), and A2 (solid line labeled "$A2$").

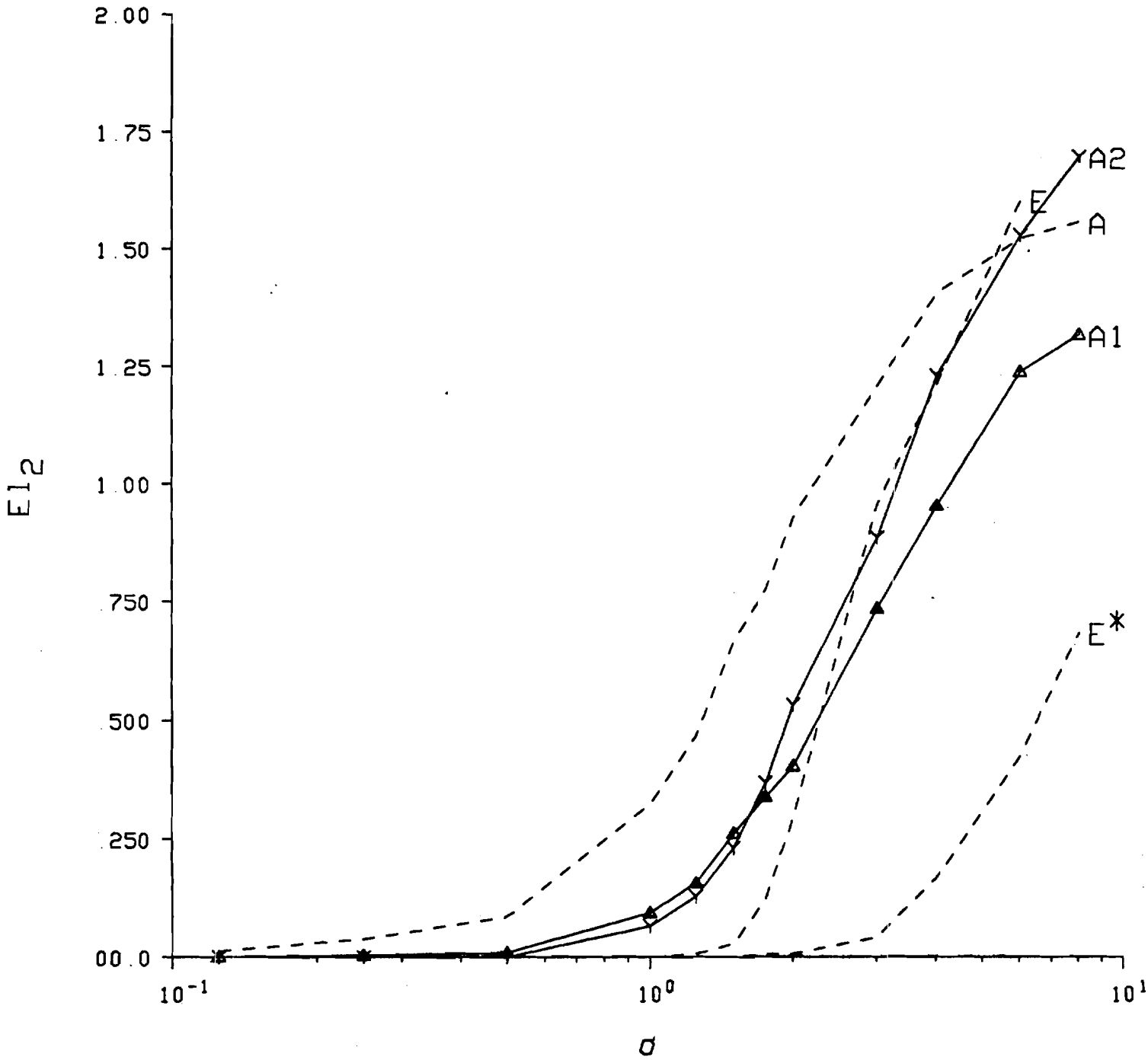


Figure 10: Estimator performance statistics: $E(l_2)$ versus σ for the three buic estimators of Section 3 (dotted lines), $A1$ (rolid line labeled "$A1$"), and A2 (solid line labeled "$A2$").

in A2 relittive to $A1$ will track the ratio of the fundamental domain volume to the unit cell volume. Therefore in more complicated space groups in higher dimensions the advantage in this sense of A2 relative to $A1$ will increase. While the dimension of the space is not the only determinant of the level of difficulty of an optimization problem, it is an important issue.

On the other hand, note that for each new space group the computation of A2 requires possibly difficult analytic calculations in order to (1) locate the critical points for the small noise asymptotics and (2) sum the contributions of the critical points to the asymptotic expansion. However, in the only three-dimensional space group that has been investigated (monoclinic $C2$), the calculations can be done and in fact are a combination of the results for Approach 2 as described in this paper and the results of Ref. [1]. In summary, it will require further analytical calculations for other space groups and numerical experimentation in order to determine the relative merits of these two estimators.

15 Asymptotics–critical point for Approach 3

In this section the system of equations defining the critical point for Approach 3 is derived corresponding to the equations of Ref. [1, Section 10]. The general solution of these equations is not known but the solution for a special case is briefly sketched.

Define: the Lagrangian $\mathcal{L}$ by

$$\mathcal{L} = \beta H_\lambda + tC_{\text{fd}}.$$

Taking derivatives with respect to r and t and setting them equal to zero gives the following system of equations for the stationary points p, τ:

$$\begin{aligned} 0 &= (\partial_t \mathcal{L})(\rho, \tau) \\ &= \rho^T \operatorname{diag}(v_0, v_1, \ldots, v_{\frac{L}{2}-1})\rho + 2\rho^T v v^T \rho - \frac{1}{2}\rho_0 - v^T \rho \\ 0 &= (\partial_{r_0} \mathcal{L})(\rho, \tau) \\ &= -b_{0,1b} - 2(b_{0,2a} + b_{0,2b})\rho_0 + 4b_{0,4a}\rho_0^3 + \tau[v_0(2\rho_0 + 4v^T\rho - 1) - \frac{1}{2}] \end{aligned}$$

$$
\begin{aligned}
0 &= (\partial_{r_k}\mathcal{L})(\rho,\tau) \\
&= -2(b_{k,2a}+b_{k,2b})\rho_k + 4b_{k,4a}\rho_k^3 + \tau v_k(2\rho_k + 4v^T\rho - 1) \quad k \in K_L^+.
\end{aligned}
$$

The second order condition, specifically, $y^T Q(\rho,\tau)y \geq 0$ for all y $\in M(\rho)$, involves the subspace

$$
\begin{aligned}
M(\rho) &= \{y : \nabla C_{fd}(\rho)^T y = 0\} \\
&= \{y : -\frac{1}{2}y_0 + \frac{1}{L}\left(\rho_0 y_0 + 2\sum_{k=1}^{\frac{L-1}{2}}\rho_k y_k\right) + (4v^T\rho - 1)\frac{1}{2L}\left(y_0 + 2\sum_{k=0}^{\frac{L-1}{2}} y_k\right) = 0\}
\end{aligned}
$$

and the Hessian matrix

$$
\begin{aligned}
Q(\rho,\tau) &= (\nabla^2\mathcal{L})(\rho,\tau) \\
&= \mathrm{diag}(-2(b_{0,2a}+b_{0,2b}) + 12b_{0,4a}\rho_0^2 + \tau\frac{1}{L}, \ldots, -2(b_{k,2a}+b_{k,2b}) + 12b_{k,4a}\rho_k^2 + \tau\frac{2}{L}, \ldots) \\
&\quad + \tau 4vv^T.
\end{aligned}
$$

Approximate numerical results seemed to indicate that solutions of the gradient equations often hacl one or more ρ components that were roughly three orders of magnitude smaller than the remaining ρ components. It seemed possible that if the numerical results were exact then these components would be exactly zero. Recall that exactly zero components can occur in the original critical point problem solved in Refs. [1, 2].

Therefore the following special case seemed of interest. Assume that there are one or more solutions of the gradient equations in which there exists a $k \in K_L^+$ such that $\rho_k = 0$. (Even if such solutions exist, they may not correspond to minima let alone global minima of the opticnization problem). It turns out that having even one such ρ_k greatly simplifies the solution of the gradient equations.

Consider the k gradient equation. Since $\rho_k = 0$ this equation simplifies to

$$
\tau v_k(4v^T\rho - 1) = 0.
$$

Therefore, either $\tau = 0$ or $4v^T p = 1$.

Assume $\tau = 0$. The k = 0 gradient equation simplifies to

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

which has three solutions at least one of which is guaranteed to be real. The $k \in K_L^+ - \{\dot{k}\}$ gradient equations simply to

$$0 = -2(b_{0,2a} + b_{0,2b})\rho_k + 4b_{0,4a}\rho_k^3$$

which has the three solutions

$$\rho_k = 0, \pm\sqrt{\frac{b_{k,2a} + b_{k,2b}}{2b_{k,4a}}}$$

of which at least one is guaranteed to be real. It seems highly unlikely that some choice among these finite set of ρ solutions, each of which depends on the data, will satisfy the constraint equation which does not depend on the data. Therefore it seems highly unlikely that $\tau = 0$ will ever occur in practice.

Now assume that $\tau \neq 0$. Then $4v^T\rho - 1 = 0$. This dramatically simplifies, and especially uncouples, the stationary point equations. Specifically, the equation $0 = (\partial_t \mathcal{L})(\rho, \tau)$ becomes

$$0 = \rho^T \operatorname{diag}(v_0, v_1, \ldots, v_{\frac{L-1}{2}})\rho - \frac{1}{2}\rho_0 - \frac{1}{8},$$

the equation $0 = (\partial_{r_0} \mathcal{L})(\rho, \tau)$ becomes

$$0 = -b_{0,1b} - 2(b_{0,2a} + b_{0,2b})\rho_0 + 4b_{0,4a}\rho_0^3 + \tau(2v_0\rho_0 - \frac{1}{2}),$$

the equation $0 = (\partial_{r_k} \mathcal{L})(\rho, \tau)$ becomes

$$0 = -2(b_{k,2a} + b_{k,2b})\rho_k + 4b_{k,4a}\rho_k^3 + \tau 2v_k\rho_k,$$

and the subspace $M(\rho)$ becomes

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

The Hessian matrix $Q(\rho, \tau)$ does not simplify.

These equations are sufficiently simple that the techniques of Ref. [1, Section 10] can be used to compute an analytic solution. However, when the solutions to many problems are carefully computed by numerical methods (see Section 17), it seems that the crucial assumption, that there exists a k such that $\rho_k = 0$ which implies that $4v^T p - 1 = 0$, is often violated. Specifically, Section 17 contains a plot (Figure 12) of $4v^T\rho - 1 = 0$ for 1000 different problems which shows that for some numerically-obtained p, $4v^T\rho - 1 = 0$ is far from zero. Therefore the analytical solution is omitted. Recall that the critical point is the global minimum. Therefore, though the problem seems difficult, further progress in this area is desirable because it is difficult to compute global solutions with purely numerical methods. This completes Step 8 for Approach 3.

16 Asymptotics–formulae for Approach 3

In this section the asymptotic formulae for Approach 3 are presented. This section closely parallels the corresponding section for Approach 2 (Section 13) but is simpler because there is only a single critical point.

The normalizer Z of the probability density is

$$\begin{aligned} Z &= \int g_Z(r)\delta(C_{\text{fd}}(r))e^{-\lambda\beta H_\lambda(r)}\mathrm{d}r \\ &= \int g_Z(F_\rho(\bar{r}))e^{-\lambda\beta H_\lambda(F_\rho(\bar{r}))}\mathrm{d}\bar{r} \\ &\approx N^{-1}(\lambda L_{\rho,\text{fd}})\exp(-\lambda\beta H_\lambda(\rho))g_Z(\rho) \end{aligned}$$

where the first transformation stems from integrating the r_0 variable and the second from taking the zeroth order term in Eq. 84 since g_Z never vanishes.

Next consider $E(\Phi_k|y)$ when $k \in K_L^+ - \bar{A}_{\rho'}$. (Note that $k = 0$ is *not* in thir set). Therefore, $\rho_k \neq 0$ which implies that g_k is not zero at the critical point and therefore only the zeroth

order term of Eq. 84 is required with the result that

$$ZE(\Phi_k|y) \approx N^{-1}(\lambda L_{\rho,\mathrm{fd}})\exp(-\lambda\beta H_\lambda(\rho))g_k(\rho)$$

which implies that

$$\begin{aligned} E(\Phi_k|y) &\approx \frac{g_k(\rho)}{gz(\rho)} \\ &= \rho_k \end{aligned}$$

For $E(\Phi_0|y)$, i.e., the k = 0 case, assume as before that $\rho_k \neq 0$. Then only the zeroth order term of Eq. 84 is required with the result that

$$E(\Phi_0|y) \approx \rho_0.$$

Finally, consider $E(\Phi_k|y)$ when $k \in \bar{A}_\rho$ which implies that $\rho_k = 0$ so it is necessary to compute the second order terms in Eq. 84. As in Section 13, the terms $J_{1,a}$ and $J_{1,b}$ of Eq. 84 are zero. However, neither $J_{1,c}$ (Appendix 24 Eq. 75) nor $J_{1,d}$ (Appendix 25 Eq. 79) are zero. Application of Eq. 84 and division by Z give the result

$$E(\Phi_k|y) \approx \frac{1}{\lambda}\frac{J_{1,c}+J_{1,d}}{gz}$$

The formulae for $J_{1,c}$ and $J_{1,d}$ are quite complicated and are in the appendices. There are two important features:

1. Though the formulae are complicated, the computation required to implement the formulae is linear in the size of the lattice and is therefore practical.

2. The ratio $J_{1,c}/gz$ is independent of the symmetry breaking function Ψ but $J_{1,d}/gz$ is dependent on Ψ_0 and Ψ_k. This dependence is the only dependence present in Approach 3.

Once approximations to $E(\Phi_k|y)$ are computed, the final estimate $\hat{\phi}_n$ of the field ϕ_n is computed, exactly as in Approach 2 as is described at the close of Section 13. This completes Step 11 for Approach 3.

17 Numerical results–Approach 3

In this section the performance of the estimator based on Approach 3 is presented. First the method of computing the location of the critical point for the small noise asymptotics must be described.

The location of the critical point for the small noise asymptotics is determined by a nonlinearly-constrained nonlinear optimization problem where both the constraint and the objective function are polynomials. In order for a correct asymptotic calculation, the critical point must be the global minimum of this optimization problem. Two methods are used for locating the critical point:

1. a homotopy continuation method as implemented in Ref. [15] and

2. a successive quadratic programming method using gradients as implemented in IMSL Edition 10.0 subroutine N2ONG (a special case of subroutine **NCONG**) documented in Ref. [16, Section 8.4, pp. 903-9081.

The homotopy method works on the system of polynomial equations that determine the stationary points. It is guaranteed in theory to compute all roots of the system, in particular including all minima and maxima. Then the value of the objective function evaluated at the stationary points is compared in order to determine the critical point. In exchange for the guarantee of a complete set of roots, the computational cost is high since each root requires the integration of a differential equation. Therefore this method it not of practical use in realistic problems for the crystallography application. However, it provides a computational fix to the fact that it is not possible to compute the critical point analytically and thereby allows estimator performance to be separated from the performance of a numerical algorithm for the location of the critical point. In terms of size, the problem of Section 3 represents the upper limit of practicality for this method and in fact, given the available computer resources, it is not possible to use this method to compute the estimator performance via

Monte Carlo methods for the problem of Section 3. Therefore results using this method are not presented.

The successive quadratic programming method works on the optimization problem directly. It does not guarantee convergence to the global minimum. In order to deal with this difficulty, multiple initial conditions are used. An important issue is how many initial conditions of what type are required in order to have a reasonably high probability of reaching the global minimum.

In the work reported here, only randomly chosen initial conditions are considered. The distribution of the initial conditions takes advantage of the known eigenvector and eigenvalue structure of the spherical-model constraint quadratic form $C_{\mathrm{fd}}(r) = (r+d_{\mathrm{fd}})^T \Sigma_{\mathrm{fd}} (r+d_{\mathrm{fd}}) - \frac{L+1}{8}$ as computed in Section 8. Specifically, define $\chi_{\min} = \min(\chi, \chi_+, \chi_-)$. It would be desirable to sample uniformly over the set $\{\mathbf{r} : (\mathbf{r}+d_{\mathrm{fd}})^T \Sigma_{\mathrm{fd}} (r+d_{\mathrm{fd}}) = \frac{L+1}{8}\}$ but this is difficult so instead sample uniformly over the larger set $\{\mathbf{r} : -\sqrt{\frac{L+1}{8\chi_{\min}}} \le r_i + d_{\mathrm{fd},i} \le \sqrt{\frac{L+1}{8\chi_{\min}}} \ \forall i\}$ which is a larger set since $\{\mathbf{r} : (\mathbf{r}+d_{\mathrm{fd}})^T \Sigma_{\mathrm{fd}} (r+d_{\mathrm{fd}}) = \frac{L+1}{8}\} \subset \{\mathbf{r} : (\mathbf{r}+d_{\mathrm{fd}})^T \Sigma_{\mathrm{fd}} (r+d_{\mathrm{fd}}) \le \frac{L+1}{8}\} \subset \{r : \|r+d_{\mathrm{fd}}\|_2^2 \chi_{\min} \le \frac{L+1}{8}\} \subset \{r : |r_i + d_{\mathrm{fd},i}|^2 \chi_{\min} \le \frac{L+1}{8} \ \forall i\} = \{\mathbf{r} : -\sqrt{\frac{L+1}{8\chi_{\min}}} \le r_i + d_{\mathrm{fd},i} \le \sqrt{\frac{L+1}{8\chi_{\min}}} \ \forall\}$. Uniform sampling over the final set can be achieved by taking vectors with components that are independent pseudorandom variables where the i th variable in a vector has a uniform distribution over the interval $[-\sqrt{\frac{L+1}{8\chi_{\min}}} - d_{\mathrm{fd},i}, \sqrt{\frac{L+1}{8\chi_{\min}}} - d_{\mathrm{fd},i})$. In view of this collection of set inclusions, more sophisticated methods of sampling are obviously possible.

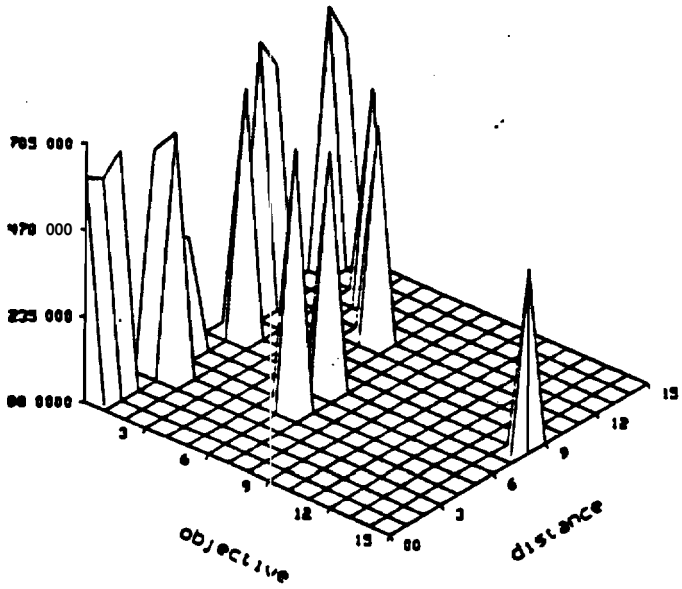
The second issue is the number of such initial conditions that are required. A rough idea of this number is determined by experimentation. For two examples the position ρ and values βH of the local minima starting from 10000 initial conditions, the position ρ^* and value βH^* of the minimum of the 10000 local minima, and the differences $\|\rho - \rho^*\|_2^2$ and $|\beta H - \beta H^*|$ are computed. A two-dimensional histogram (essentially an estimate of the joint probability density function) of the differences $\|\rho - \rho^*\|_2^2$ and $|\beta H - \beta H^*|$ is computed and portions are displayed as surface plots.

The first example problem is one for which the observation noise standard deviation is

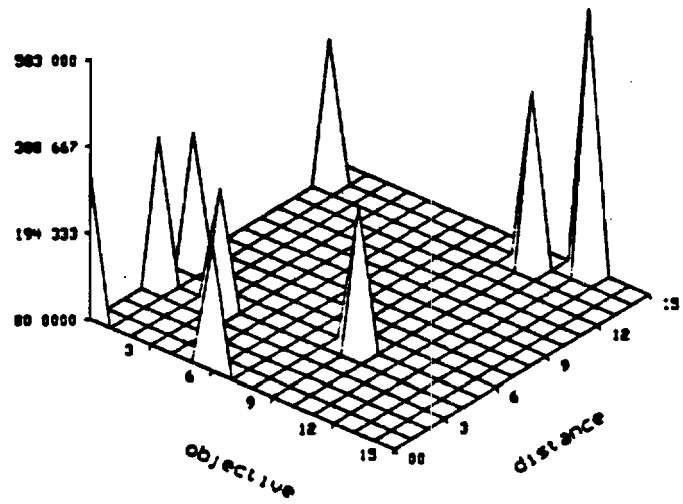
$\sigma = 2$. Calculations are done in double precision using DN2ONG rather than N2ONG. For this problem, the minimum objective function local minima is -168.6 and the maximum is -162.1 and the minimum location difference between some other ρ and ρ^* (i.e., $\|\rho - \rho^*\|_2^2$) is 2.044×10^{-8} and the maximum is 125.6. A 16 by 16 bin histogram displaying the entire range of observed values for $\|\rho - \rho^*\|_2^2$ (roughly 0 to 125.6) and $|\beta H - \beta H^*|$ (roughly 0 to 6.48) is shown in Figure 11a. There are more than 10 peaks corresponding to local minima that are found starting from multiple initial conditions. A 16 by 16 bin histogram displaying the entire range of $\|\rho - \rho^*\|_2^2$ but only 1/64 of the range of $|\beta H - \beta H^*|$ is shown in Figure 11b. At this higher scale, many of the peaks of Figure 11a are subdivided, and there are 4 local minima that, at this scale, achieve the lowest value of βH. The (0,0) bin contains 327 initial condition, trials. Finally a 16 by 16 bin histogram displaying the entire range of $\|\rho - \rho^*\|_2^2$ but only 1/4096 of the range of $|\beta H - \beta H^*|$ is shown in Figure 11c. Only two local minima appear and only one of them [the (0,0) bin] achieves the lowest value of βH. The (0,0) bin continues to contain 327 initial condition trials. These 327 trials have the 327 lowest values of βH among the 10000 trials. Within this class, the norm squared ($\|\rho\|_2^2$) of the locations ρ have the following sample statistics: the minimum is 57.456, the maximum is 57.457, the sample mean is 57.456, and the sample variance is 8.0×10^{-9}. Within this class, the norm squared of pairwise differences ($\|\rho - \rho'\|_2^2$) of the locations ρ has the following sample statistics: the minimum is 1.1×10^{-11}, the maximum is 1.3×10^{-5}, the sample mean is 1.7×10^{-7} and, the sample variance is 3.1×10^{-13}. Therefore the differences among the results of trials in this class are of the order of numerical errors.

Consider a Bernoulli process model of the sequence of trials where a successful trial is defined to be a trial that locates the global minimum. Then an estimate of the probability of success on an individual trial is p = 327/10000, and the probability of no successes in n trials is $(1 - p)^n$. By taking 2000 trials for problems with $\sigma = 2$, the probability of no successes, (i.e., failure to locate the global minimum) as computed by this model is small.

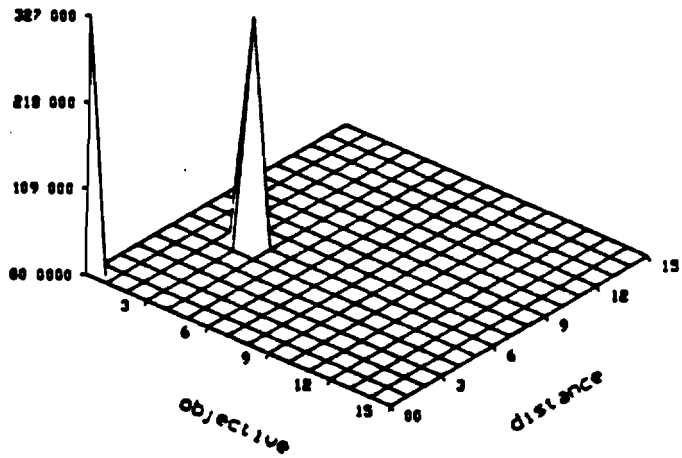
The second example problem is for $\sigma = .5$. For this problem a larger number, specifically



(a)



(b)



(c)

Figure 11: Two dimensional histograms for $\|\rho - \rho^*\|_2^2$ and $|\beta H - \beta H^*|$ shown as surface plots. Each plot shows the entire observed range of values for $\|\rho - \rho^*\|_2^2$. Different fractions starting from the origin of the entire observed range of values for $|\beta H - \beta H^*|$ are displayed: Part (a) displays the entire range, Part (b) displays 1/64, and Part (c) displays 1/4096.

4000, initial conditions seem worthwhile based on histograms that are not shown. Therefore, for all problems at $\sigma = .5$, 4000 initial conditions are tried.

Having settled on this method for computing the critical points in the estimator using Approach 3, the applicability of the special-case analytical solution described in Section 15 can be considered. Recall that the special case is defined by the assumption that there exists a k such that $\rho_k = 0$. This assumption then implies that $4v^T\rho - 1 = 0$. Based on the previously described numerical method for computing the critical points, Figure 12 shows a plot of the fraction of $a = 2$ data sets for which $4v^T\rho - 1$ is less than a given value. This is essentially an estimate of the cumulative probability distribution function of $4v^Tp - 1$. Clearly, there are many data sets for which $4v^Tp - 1$ is far from 0. Whether these represent failures of the numerical algorithm for locating the critical points or true cases of $4v^Tp - 1$ far from 0 is not clear. However, for this initial investigation, it seemed more interesting to focus on the performance of the complete estimator rather than the critical point determination especially since this estimator has a novel featurt–little need for and opportunity for symmetry breaking optimization.

The ,performanceof the estimator based on Approach 3, denoted "*A3*", can now be computed and compared with five alternative estimators. The problem and three of alternative estimators–the basic estimators *E-P*1̄, *E-P*1, and A–are discussed in Section 3. The fourth alternative estimator, *A*1 based on Approach 1, is discussed in Sections 4 and 5. The fifth alternative estimator, A2 based on Approach 2, is discussed in Sections 6, 7, 8, 9, 10, 11, 12, 13, and 14.

The primary results are shown in Figures 13 and 14. Parameters for the three basic estimators are described in Section 3. Parameters for *A*1 (A2) are described in Section 5 (14). Estimator A3 is matched to the synthetic data. Because the critical point locations are determined numerically, there are no identically zero components. Therefore the zeroth order term in the asymptotic expansion is nonzero, the second order term is not computed, the symmetry breaking kernel ψ does not influencethe final solution, and symmetry breaking

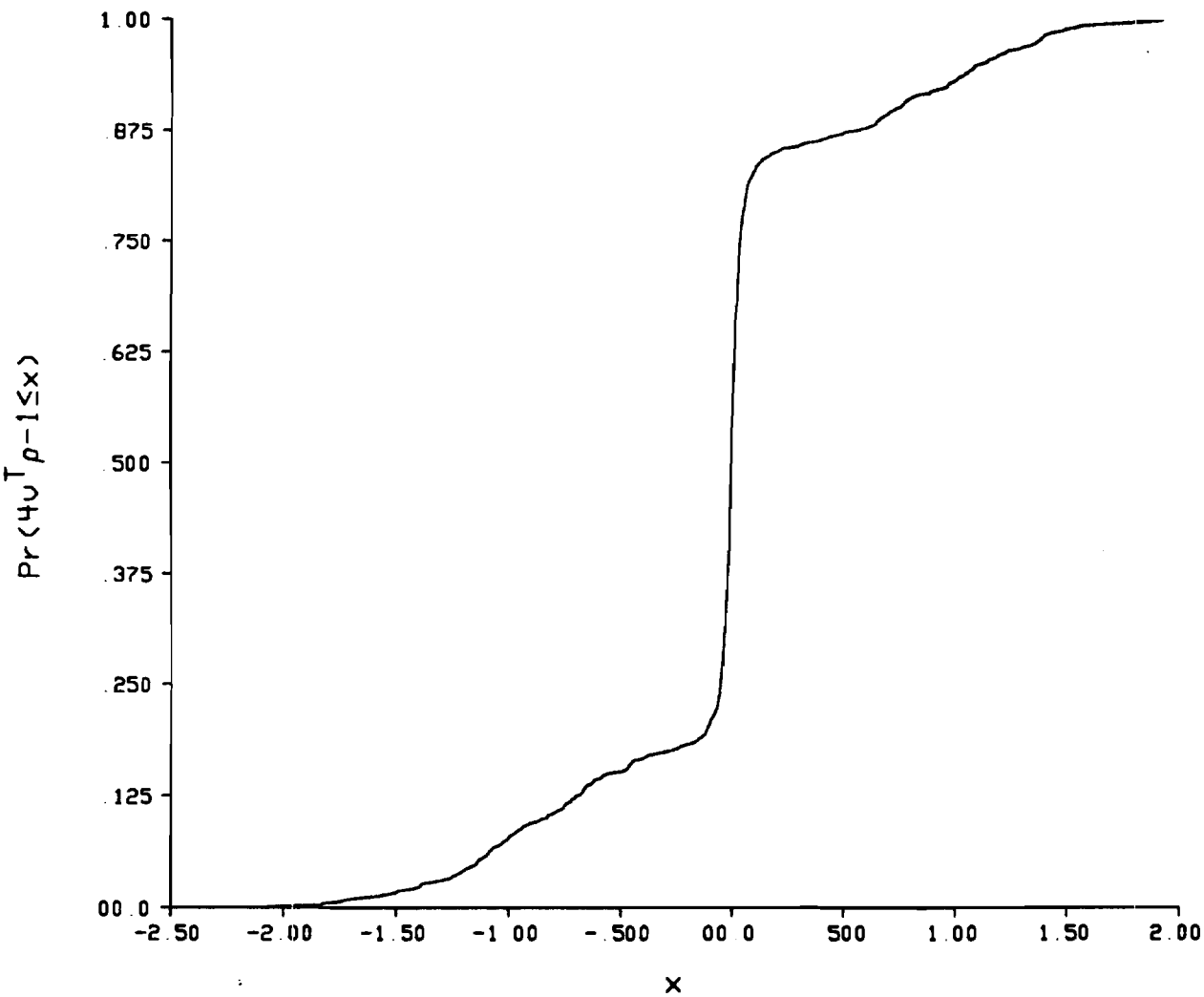


Figure 1.2: Estimate of the cumulative probability distribution function of $4v^T\rho - 1$.

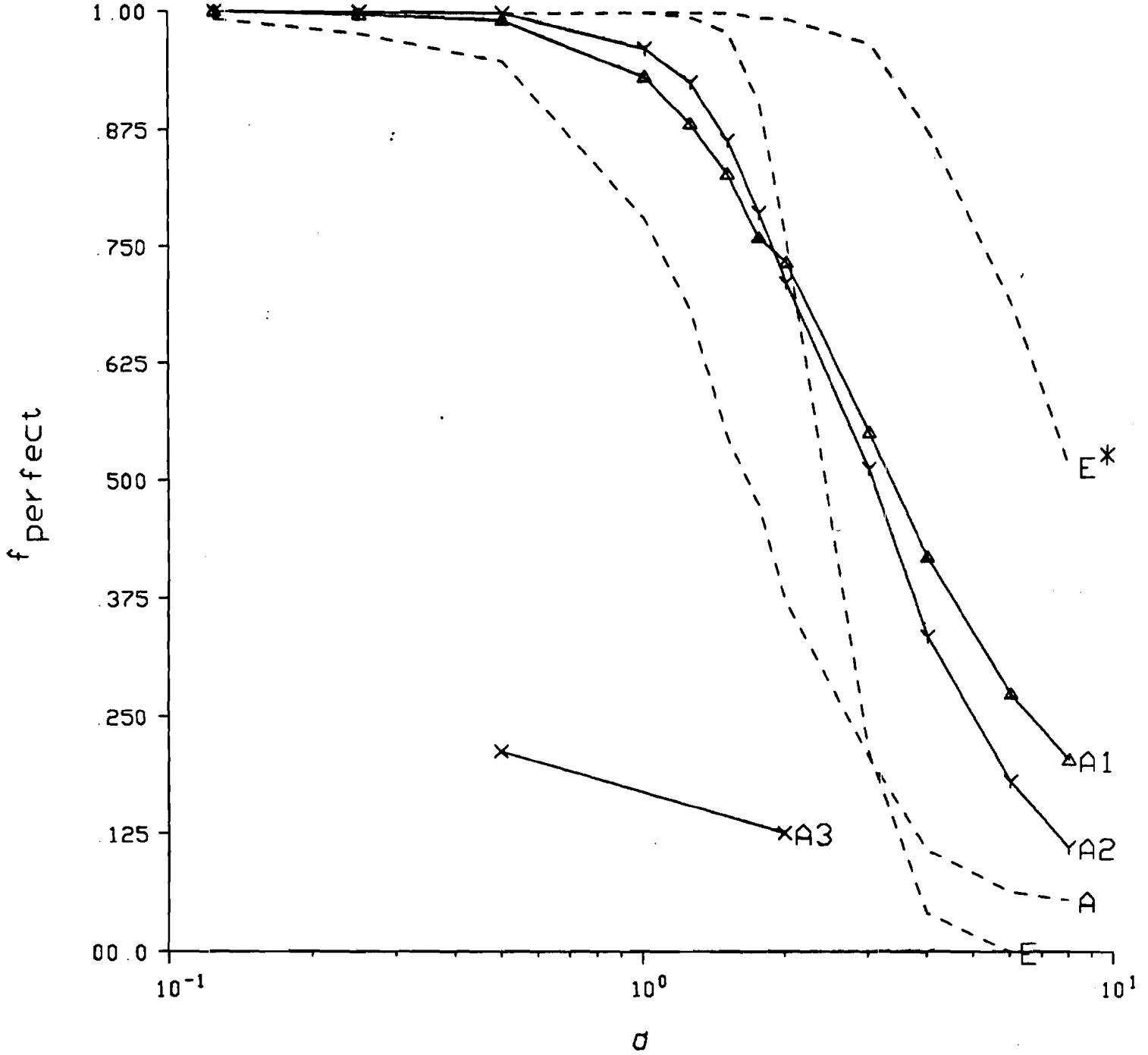


Figure 13: Estimator performance statistics: $f_{perfect}$ versus σ for the throe basic estimators of Section 3 (dotted lines), *A*1 (solid line labeled "*A*1"), A2 (solid line labeled "*A*2"), and A3 (solid line labeled "*A*3").

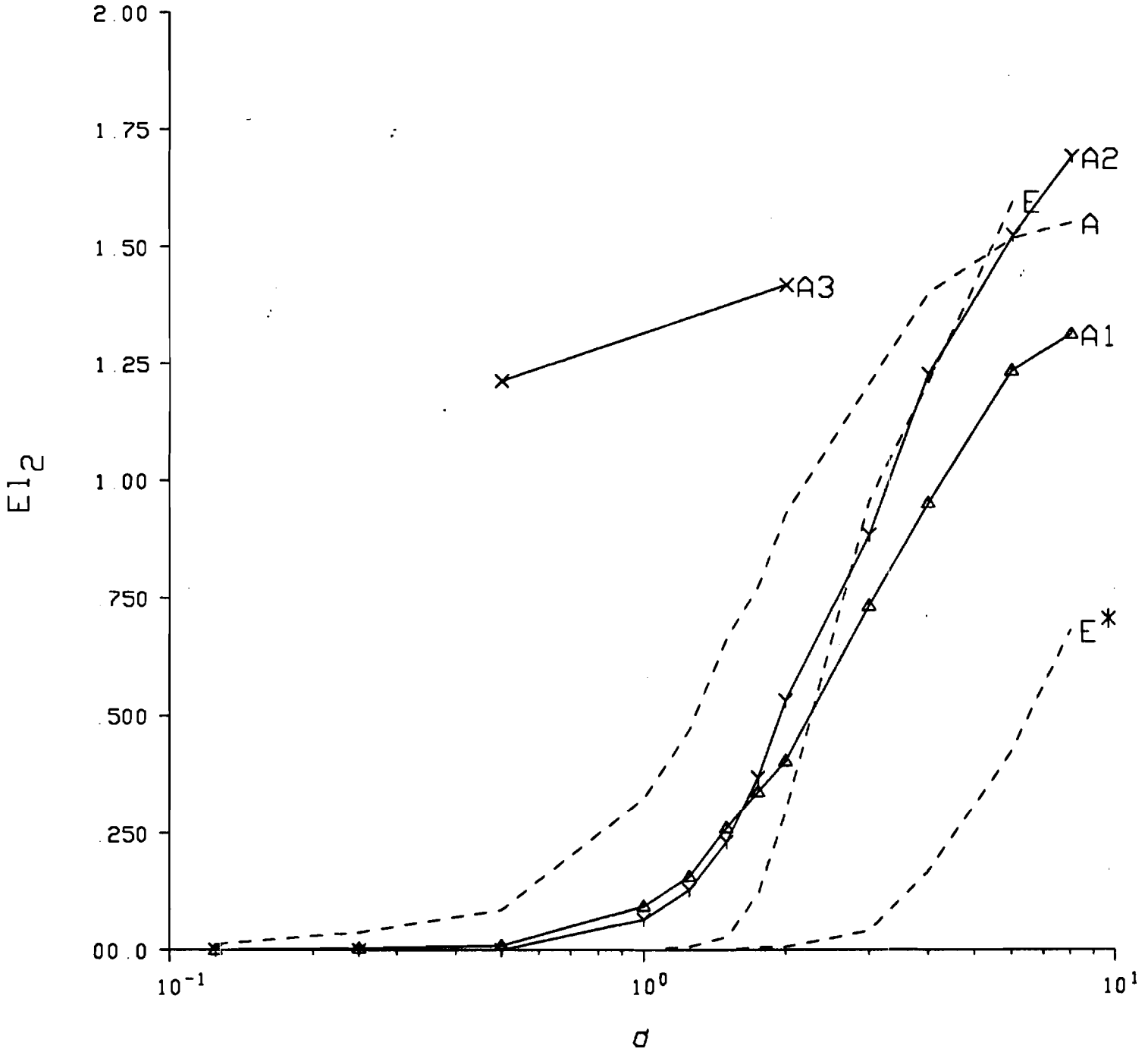


Figure 14: Estimator performance statistics: $E(l_2)$ versus σ for the three basic estimators of Section 3 (dotted lines), $A1$ (solid line labeled "$A1$"), A2 (solid line labeled "$A2$"), and A3 (solid line labeled "$A3$").

optimization is turned off. This is discussed in Section 16. Estimator A3 has in addition $\chi = 0.5$, $\lambda = 1.0$, and $\beta = 1.0$.

The performance of estimator A3 is consistently poor. For that reason it is only computed for the two values of σ indicated in Figures 13 and 14. Since estimator A3 is the only estimator studied which lacks significant symmetry breaking and the data adaptation that symmetry breaking optimization allows, it is natural to attribute the poor performance to this reason. Therefore, one of the main conclusions of the numerical work described in this paper is the importance of data adaptation in achieving good estimator performance. This is not an unexpected result since it is only through data adaptation that the spherical model approximation is ameliorated. However, there are two areas where improvements might lead to better performance. First, the assumptions behind the asymptotic calculations require that the critical point be the global minimum. The numerical techniques used here may not reliably locate the global rather than a local minimum. Therefore, further development of solution techniques (especially analytical techniques) is desirable. Second, if Figure 12 is an accurate representation of the local minima structure of typical examples, then there are multiple local minima lying close in value to the global minimum. In the $\lambda \to \infty$ limit the contributions from these local minima are negligible. However, at the $\lambda = 1$ value used in computing the estimators, the contributions may be significant. Therefore asymptotic formulae and computer software that include contributionr from multiple low-lying local minima are desirable. The algorithm to include contributionr from multiple local minima is not trivial when the minima are located numerically because the same local minimum will typically be located many times with slightly different locations and the clustering of the locations must be recognized in order to avoid including the single local minimum as several distinct local minima.

18 Discussion and Future Directions

In this paper three methods are presented for incorporating symmetry constraints into a signal reconstruction problem based on Fourier transform magnitude measurements that was introduced in Ref. [1]. As shown in Figures 3 and 4, the performance gain possible through exploitation of the symmetry is large. Approach 1 in its most powerful form (i.e., $A1$) encourages symmetric solutions by adding an appropriate term to the symmetry-breaking optimization criteria and then, in a postprocessing phase, enforces a symmetric solution by averaging around orbits. Approachs 2 and 3 (i.e., A2 and A3) apply the symmetry as a hard constraint in the basic signal reconstruction algorithm. They differ in the order in which two noncommuting nonlinear operations are performed. In Approach 2 the spherical model is applied before the symmetry constraint while in Approach 3 the reverse order is used.

In terrns of performance, Approaches 1 and 2 have a clear advantage over Approach 3. In fact, either of Approaches 1 or 2 outperforms the exact estimator E-$P1$ which, however, is not aware of the presence of symmetry. On the other hand, Approach 3 is outperformed by estimator A of Refs. [1, 2] which is not aware of the presence of symmetry. The inferior performance of Approach 3 is likely due to the limited opportunities in Approach 3 for data-adaptive symmetry breaking optimization which is used to ameliorate the effects of the spherical model approximation in all of the other approximate estinnators. However, two aspects of the implementation of the small noise asymptotics may play a role: First, the numerical rather than analytical calculation of the critical point location may fail to locate the global minimum and second the asymptotic formulae which include only the global minimum may be misleading when used at $\lambda = 1$ because of the presence of several low-lying local minima.

In terms of computation, Approach 2 has an important advantage over Approach 1 because, in the case of $P\bar{1}$, the symmetry breaking optimization occurs in a space of one half the dimension. In more complicated space groups for which the fundamental domain is

a smaller fraction of the unit cell than the fraction 1/2 that occurs in the case of $P\bar{1}$, it is anticipated that this advantage will increase. Approach 3, so long as the critical point must be located numerically, is at a disadvantage relative to both Approaches 1 and 2.

The number of space groups is quite large [4, p. 12]: in d-dimensions there are 2 for d = 1, 17 for d = 2, and 230 for the d = 3 case of primary interest in the crystallography application. Therefore the amount of analysis required to apply an approach to a particular space group is a concern. (Note, however, that especially for large-molecule structures, a relatively small subset of the space groups accounts for the bulk of the structures). Approach 1 requires the least work since tabulations of the orbits for each space group are readily available. Approach 2 requires an intermediate amount of work. For example, in the case of $P\bar{1}$ in one dimension, the critical point location from Ref. [1] is used essentially unchanged but the asymptotic evaluation of the conditional mean integrals differs. Furthermore, in the three dimensional monoclinic $C2$ space group, both the critical point location and the asymptotics require modifications which, however, can be done with a combination of the ideas from Ref [1] and the present paper. Approach 3 requires the most work since even in the case of the one dimensional $P\bar{1}$ space group the calculation of the critical point must be redone and to date only a special case has been solved analytically.

In summary, Approaches 1 and 2 appear to be rather equally matched with both having strong and weak points. As mentioned previously, both of these approaches are currently being pursued for the three dimensional monoclinic $C2$ space group for which experimental data for the crystallography application is available.

Finally, throughout this paper the symmetries considered have all been space group symmetries. However, in certain crystallographic problems, there is a second form of symmetry called noncrystallographic symmetry. This occurs when the object making up the crystal has a type of symmetry itself, such as a five fold axis of rotation, that is forbidden to any space group symmetry [6, Section 8]. It should be possible to use the ideas of Approach 1--postprocessing by averaging and modification of the symmetry breaking objective

function–to include this type of symmetry in both Approachs 1 and 2. Approach 3, because of the weak dependence of the answer on the symmetry breaking kernel. ψ, would be less appropriate.

19 Acknowledgements

I would like to thank Alexander Morgan for the software used in Section 17' to compute roots of systems of polynomials using continuation methods. This work was supported by U. S. National Science Foundation grant MIP-9110919, a Whirlpool Faculty Fellowship, and the School of Electrical Engineering, Purdue University.

20 Appendix–definitions of $a_{k,j}$ and $b_{k,ns}$

This appendix defines the $a_{k,j}$ constants (see Section 8) and the $b_{k,ns}$ constants (see Section 11). 'The $a_{k,j}$ are

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

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

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

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

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

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

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

$$b_{k,2a} = \begin{cases} \frac{\beta}{\partial_0^2} y_0 & \text{if } k = 0 \\ \frac{\beta}{\partial_k^2} y_k + \frac{\beta}{\partial_{L-k}^2} y_{L-k} & \text{if } k \neq 0 \end{cases}$$

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

$$b_{k,4a} = \begin{cases} \frac{\beta}{2\partial_0^2} & \text{if } k = 0 \\ \frac{\beta}{2\partial_k^2} + \frac{\beta}{2\partial_{L-k}^2} & \text{if } k \neq 0 \end{cases}.$$

For Problem 2 the definitions are

$$b_{k,0a} = \begin{cases} \frac{\beta}{2\partial_0^2} y_0^2 & \text{if } k = 0 \\ \frac{\beta}{2\partial_k^2} y_k^2 + \frac{\beta}{2\partial_{L-k}^2} y_{L-k}^2 & \text{if } k \neq 0 \end{cases}$$

$$b_{k,1b} = \begin{cases} -\beta\chi w_1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

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

$$b_{k,2a} = \begin{cases} \frac{\beta}{\partial_0^2} y_0 & \text{if } k = 0 \\ \frac{\beta}{\partial_k^2} y_k + \frac{\beta}{\partial_{L-k}^2} y_{L-k} & \text{if } k \neq 0 \end{cases}$$

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

$$b_{k,4a} = \begin{cases} \frac{\beta}{2\partial_0^2} & \text{if } k = 0 \\ \frac{\beta}{2\partial_k^2} + \frac{\beta}{2\partial_{L-k}^2} & \text{if } k \neq 0 \end{cases}.$$

For both Problems 1 and 2 note that the only dependence on Ψ_k is in $b_{k,1c}$.

21 Appendix–invariance of $\det L_{\rho,\mathrm{uc}}$

In this appendix it is shown that $\det L_{\rho,\mathrm{uc}}$ is constant for all ρ derived from a fixed ρ' by sign changes. Consider ρ^1 and ρ^2 differing only in the sign of one component call it k. Corresponding elements of $L_{\rho^1,\mathrm{uc}}$ and $L_{\rho^2,\mathrm{uc}}$ are equal except for a possible change of sign. The signs differ only on the k th row and the k th column with the exception that the

diagonal element (k^t, k^t) has the same sign in both $L_{\rho^1,uc}$ and $L_{\rho^2,uc}$. Therefore,

$$L_{\rho^1,uc} = \text{diag}(1,\ldots,1,-1,1,\ldots,1)L_{\rho^2,uc}\,\text{diag}(1,\ldots,1,-1,1,\ldots,1)$$

where the "-1" is in the k^t position of each diagonal matrix. Therefore, $\det L_{\rho^1,uc} = \det L_{\rho^2,uc}$ since $\det(AB) = \det(A)\det(B)$ and $\det \text{diag}(1,\ldots,1,-1,1,\ldots,1) = -1$. Since any ρ can be reached from ρ^t by a sequence of ρ^k where ρ^k and ρ^{k+1} differ only by one sign, $\det L_\rho = \det L_{\rho^t}$ for any ρ.

22 Appendix–Derivatives of $\beta H_\lambda(F_\rho(e))$ for Approaches 2 and 3

In this appendix the zeroth through third order partial derivatives of $D_\rho(\bar{r}) = \beta H_\lambda(F_\rho(\bar{r}))$ are computed and the results are evaluated at the critical point ρ. In the! formulae for the asymptotic evaluation of integrals such as Eqs. 10, 11, and 12 it is necessary to have formulae for the inverse of the Hessian matrix of $D_\rho(\bar{r})$ evaluated at the critical point $\bar{r} = \bar{\rho}$. (The Hessian matrix, denoted $\nabla^2 D_\rho(\bar{r})$, is the matrix of mixed second partial derivatives). These formulae are also computed in this appendix. Since D_ρ is a function of $\bar{r}$, these derivatives are with respect to r_i for $i \neq 0$.

Recall the definitions of η_ρ and F_ρ from Eqs. 19 and 20 and that there! are actually two η_ρ and two F_ρ functions corresponding to the two C functions C_{uc} and C_{fd}. In the remainder of this appendix, when the argument of C, H_λ and their derivatives (respectively η_ρ and its derivatives) is omitted, then $F_\rho(\bar{r}) = (\eta_\rho(\bar{r}), \bar{r})^T$ (respectively $\bar{r}$) is to be assumed.

Using the chain rule, take a derivative of Eq. 19, the definition of η_ρ, to get

$$(\partial_{r_i}\eta_\rho)(\bar{r}) = -\left[(\partial_{r_0}C)(F_\rho(\bar{r}))\right]^{-1}(\partial_{r_i}C)(F_\rho(\bar{r})). \tag{35}$$

Taking a derivative of Eq. 35 gives

$$(\partial_{r_i}\partial_{r_j}\eta_\rho) = -(\partial_{r_0}C)^{-3}(\partial_{r_0}\partial_{r_0}C)(\partial_{r_i}C)(\partial_{r_j}C)$$

$$+ (\partial_{r_0} C)^{-2} \left[(\partial_{r_i} \partial_{r_0} C)(\partial_{r_j} C) + (\partial_{r_0} \partial_{r_j} C)(\partial_{r_i} C) \right]$$
$$- (\partial_{r_0} C)^{-1} (\partial_{r_i} \partial_{r_j} C). \tag{36}$$

If the partial derivatives of C do not depend on the order in which they are taken (i.e., $(\partial_{r_i} \partial_{r_j} C) = (\partial_{r_j} \partial_{r_i} C)$ and so forth) then the partial derivatives of η_ρ are also independent of the order.

Since $F_\rho(\bar{\rho}) = \rho$, the zeroth order coefficient $D_\rho(\bar{\rho})$ is simply $D_\rho(\bar{\rho}) = \beta H_\lambda(\rho)$.

The first and second order coefficients are more complicated. The definition of D_ρ and Eqs. 35 and 36 imply that

$$(\partial_{r_j} D_\rho) = -(\partial_{r_0} \beta H_\lambda)(\partial_{r_0} C)^{-1} (\partial_{r_j} C) + (\partial_{r_j} \beta H_\lambda) \tag{37}$$

$$\begin{aligned} (\partial_{r_i} \partial_{r_j} D_\rho) &= [(\partial_{r_0} \partial_{r_0} \beta H_\lambda)(\partial_{r_i} \eta_\rho) + (\partial_{r_i} \partial_{r_0} \beta H_\lambda)] (\partial_{r_j} \eta_\rho) \\ &\quad + (\partial_{r_0} \beta H_\lambda)(\partial_{r_i} \partial_{r_j} \eta_\rho) + (\partial_{r_0} \partial_{r_j} \beta H_\lambda)(\partial_{r_i} \eta_\rho) + (\partial_{r_i} \partial_{r_j} \beta H_\lambda) \qquad (38) \\ &= \left[(\partial_{r_0} \partial_{r_0} \beta H_\lambda) - \frac{(\partial_{r_0} \beta H_\lambda)}{(\partial_{r_0} C)} (\partial_{r_0} \partial_{r_0} C) \right] (\partial_{r_0} C)^{-1} (\partial_{r_i} C)(\partial_{r_0} C)^{-1} (\partial_{r_j} C) \\ &\quad - \left[(\partial_{r_i} \partial_{r_0} \beta H_\lambda) - \frac{(\partial_{r_0} \beta H_\lambda)}{(\partial_{r_0} C)} (\partial_{r_i} \partial_{r_0} C) \right] (\partial_{r_0} C)^{-1} (\partial_{r_j} C) \\ &\quad - \left[(\partial_{r_0} \partial_{r_j} \beta H_\lambda) - \frac{(\partial_{r_0} \beta H_\lambda)}{(\partial_{r_0} C)} (\partial_{r_0} \partial_{r_j} C) \right] (\partial_{r_0} C)^{-1} (\partial_{r_i} C) \\ &\quad + \left[(\partial_{r_i} \partial_{r_j} \beta H_\lambda) - \frac{(\partial_{r_0} \beta H_\lambda)}{(\partial_{r_0} C)} (\partial_{r_i} \partial_{r_j} C) \right]. \qquad (39) \end{aligned}$$

Eqs. 37 and 39 are valid for any argument $\bar{r}$. Now specialize to the case when $\bar{r} = \bar{\rho}$. The first order conditions for a local minimum of $\beta H_\lambda(r)$ subject to $C(r) = 0$ include $(\partial_{r_i} \beta H_\lambda)(\rho) = -\tau(\partial_{r_i} C)(\rho)$ for all $i \in K_L$ where τ is the Lagrange multiplier. Using this condition in Eqs. 37 and 39 gives simpler results, valid only at the critical point p, which are

$$\begin{aligned} (\partial_{r_j} D_\rho)(\bar{\rho}) &= \tau(\partial_{r_0} C)(\partial_{r_0} C)^{-1} (\partial_{r_j} C) - \tau(\partial_{r_j} C) \\ &= 0 \qquad (40) \end{aligned}$$

and (requiring only the k = 0 case of the first order condition)

$$(\partial_{r_i} \partial_{r_j} D_\rho)(\bar{\rho}) = [(\partial_{r_0} \partial_{r_0} \beta H_\lambda) + \tau(\partial_{r_0} \partial_{r_0} C)] (\partial_{r_0} C)^{-1} (\partial_{r_i} C)(\partial_{r_0} C)^{-1} (\partial_{r_j} C)$$

$$- [(\partial_{r_i}\partial_{r_0}\beta H_\lambda) + \tau(\partial_{r_i}\partial_{r_0}C)](\partial_{r_0}C)^{-1}(\partial_{r_j}C)$$
$$- \left[(\partial_{r_0}\partial_{r_j}\beta H_\lambda) + \tau(\partial_{r_0}\partial_{r_j}C)\right](\partial_{r_0}C)^{-1}(\partial_{r_i}C)$$
$$+ \left[(\partial_{r_i}\partial_{r_j}\beta H_\lambda) + \tau(\partial_{r_i}\partial_{r_j}C)\right]. \qquad (41)$$

The function H_λ is the same for Approaches 2 and 3. Recall Eq. 21 which defines $f_k(r_k)$. Computing the partial derivatives of βH_λ and substituting into Eq. 41 gives

$$\begin{aligned}(\partial_{r_i}\partial_{r_j}D_\rho)(\bar{\rho}) &= [f_0(\rho_0) + \tau(\partial_{r_0}\partial_{r_0}C)](\partial_{r_0}C)^{-1}(\partial_{r_i}C)(\partial_{r_0}C)^{-1}(\partial_{r_j}C)\\ &\quad - \tau(\partial_{r_i}\partial_{r_0}C)(\partial_{r_0}C)^{-1}(\partial_{r_j}C)\\ &\quad - \tau(\partial_{r_0}\partial_{r_j}C)(\partial_{r_0}C)^{-1}(\partial_{r_i}C)\\ &\quad + \left[\delta_{i,j}f_i(\rho_i) + \tau(\partial_{r_i}\partial_{r_j}C)\right]. \qquad (42)\end{aligned}$$

At this point Approaches 2 and 3 diverge because they have different definitions for C. In the following paragraphs, first for Approach 2 and then for Approach 3, $\nabla^2 D_\rho(\bar{\rho})$ (the Hessian matrix evaluated at the critical point) and then $(\nabla^2 D_\rho(\bar{\rho}))^{-1}$ are computed.

The calculation for Approach 2 is simpler than for Approach 3 because all of the mixed second order partial derivatives of C_{uc} are zero. Computing the partial derivatives of C_{uc} and substituting into Eq. 42 gives the result

$$(\partial_{r_i}\partial_{r_j}D_{\rho,uc})(\bar{\rho}) = \left[f_0(\rho_0) + \tau\frac{2}{L}\right]\frac{\frac{4}{L}\rho_i\frac{4}{L}\rho_j}{\left(\frac{2}{L}\rho_0 - 1\right)^2} + \delta_{i,j}\left[f_i(\rho_i) + \tau\frac{4}{L}\right] \qquad (43)$$

valid only at the critical point p. This result can be written in the form

$$(\partial_{r_i}\partial_{r_j}D_{\rho,uc})(\bar{\rho}) = \delta_{i,j}e_{i,uc} + \rho_i'\rho_j$$

where

$$e_{i,uc} = f_i(\rho_i) + \tau\frac{4}{L} \qquad (44)$$
$$\text{and} \quad \rho_i' = \frac{f_0(\rho_0) + \tau\frac{2}{L}}{\left(\frac{2}{L}\rho_0 - 1\right)}\left(\frac{4}{L}\right)^2 \rho_i$$

for $i \in K_L^+$. Therefore $\nabla^2 D_{\rho,\mathrm{uc}}(\rho) = \mathrm{diag}(e_{i,\mathrm{uc}}) + \rho'\rho^T$ which is a diagonal matrix plus a rank 1 matrix. The **Woodbury** formula [12, p. 76] can be used to **compute** the inverse of this matrix with the result that

$$\left[\nabla^2 D_{\rho,\mathrm{uc}}(\rho)\right]^{-1} = \mathrm{diag}\left(\frac{1}{e_{i,\mathrm{uc}}}\right) - \frac{1}{\sum_{k=1}^{\frac{L-1}{2}} \frac{\rho_k^2}{e_{k,\mathrm{uc}}} + R}\gamma_{\mathrm{uc}}\gamma_{\mathrm{uc}}^T \tag{45}$$

where

$$\gamma_{\mathrm{uc}} = \left[\frac{\rho_1}{e_{1,\mathrm{uc}}}, \cdots, e_{\frac{L-1}{2},\mathrm{uc}}^{\frac{\rho_{\frac{L-1}{2}}}{2}}\right]^T \tag{46}$$

$$\text{and} \quad R = \frac{\left(\frac{2}{L}\rho_0 - 1\right)^2 e_{\frac{L-1}{2},\mathrm{uc}}}{f_0(\rho_0) + \tau\frac{2}{L}}\left(\frac{L}{4}\right)^2. \tag{47}$$

The computation of $\mathbf{V}^2 D_{\rho,\mathrm{fd}}(\rho)$ and $[\nabla^2 D_{\rho,\mathrm{fd}}(\rho)]^{-1}$ is more complicated than the computation of $\mathbf{V}^2 D_{\rho,\mathrm{uc}}(\rho)$ and $[\nabla^2 D_{\rho,\mathrm{uc}}(\rho)]^{-1}$ because the second mixed **partial** derivatives of C_{fd} are nonzero. However, the Hessian of C_{fd} **is** a diagonal matrix plus a rank 2 matrix so, as is seen in **the** following, it is again possible to use the **Woodbury** formula effectively.

Define z with components z_i for $i \in K_L^+$ by

$$\begin{aligned} z_i &= \frac{(\partial_{r_i} C_{\mathrm{fd}})(\rho)}{(\partial_{r_0} C_{\mathrm{fd}})(\rho)} \\ &= \frac{\frac{1}{L}(2\rho_k + 4v^T\rho - 1)}{\frac{1}{2L}(2\rho_0 + 4v^T\rho - 1) - \frac{1}{2}}. \end{aligned}$$

Computing the other partial derivatives of C_{fd} and substituting into Eq. 42 gives the result

$$(\partial_{r_i}\partial_{r_j} D_{\rho,\mathrm{fd}})(\bar{\rho}) = \left[f_0(\rho_0) + \tau\frac{1}{L}\right] z_i z_j + \tau\frac{1}{L^2}(z_i - 2)(z_j - 2) + \delta_{i,j}\left[f_i(\rho_i) + \tau\frac{2}{L}\right].$$

Define

$$\begin{aligned} e_{i,\mathrm{fd}} &= f_i(\rho_i) + \tau\frac{2}{L}, \\ \gamma_{i,\mathrm{fd}} &= \sqrt{f_0(\rho_0) + \tau\frac{1}{L}}\, z_i, \\ \gamma'_{i,\mathrm{fd}} &= \sqrt{\frac{\tau}{L^2}}(z_i - 2), \\ \gamma_{\mathrm{fd}} &= (\gamma_{1,\mathrm{fd}}, \ldots, \gamma_{\frac{L-1}{2},\mathrm{fd}})^T, \\ \gamma'_{\mathrm{fd}} &= (\gamma'_{1,\mathrm{fd}}, \ldots, \gamma'_{\frac{L-1}{2},\mathrm{fd}})^T, \\ \text{and} \quad \Gamma &= [\gamma_{\mathrm{fd}}, \gamma'_{\mathrm{fd}}] \in R^{(L-1)/2\times 2}. \end{aligned}$$

Then $\nabla^2 D_{\rho,\mathrm{fd}}(\rho) = \mathrm{diag}(e_{i,\mathrm{fd}}) + \Gamma\Gamma^T$ which is a diagonal matrix plus a rank 2 matrix.

Application of the Woodbury formula [12, p. 76] followed by simplification leads to the result that

$$\left[\nabla^2 D_{\rho,\mathrm{fd}}(\rho)\right]^{-1} = \mathrm{diag}\left(\frac{1}{e_{i,\mathrm{fd}}}\right) - \frac{1}{\Delta}(\alpha\zeta^T + \alpha'\zeta'^T). \tag{48}$$

where

$$\begin{aligned}
d_1 &= 1 + \sum_{k=1}^{\frac{L-1}{2}} \frac{\gamma_{k,\mathrm{fd}}^2}{e_{k,\mathrm{fd}}} \\
d_2 &= 1 + \sum_{k=1}^{\frac{L-1}{2}} \frac{\gamma'_{k,\mathrm{fd}}{}^2}{e_{k,\mathrm{fd}}} \\
d_3 &= \sum_{k=1}^{\frac{L-1}{2}} \frac{\gamma_{k,\mathrm{fd}}\gamma'_{k,\mathrm{fd}}}{e_{k,\mathrm{fd}}} \\
\Delta &= d_1 d_2 - d_3^2 \\
\zeta_k &= \frac{\gamma_{k,\mathrm{fd}}}{e_{k,\mathrm{fd}}} \\
\zeta'_k &= \frac{\gamma'_{k,\mathrm{fd}}}{e_{k,\mathrm{fd}}} \\
\alpha_k &= \zeta_k d_2 - \zeta'_k d_3 = \frac{(\gamma_{k,\mathrm{fd}} d_2 - \gamma'_{k,\mathrm{fd}} d_3)}{e_{k,\mathrm{fd}}} \\
\alpha'_k &= \zeta'_k d_1 - \zeta_k d_3 = \frac{(\gamma'_{k,\mathrm{fd}} d_1 - \gamma_{k,\mathrm{fd}} d_3)}{e_{k,\mathrm{fd}}} \\
\zeta &= (\zeta_1, \ldots, \zeta_{\frac{L-1}{2}})^T \\
\zeta' &= (\zeta'_1, \ldots, \zeta'_{\frac{L-1}{2}})^T \\
\mathrm{a} &= (\alpha_1, \ldots, \alpha_{\frac{L-1}{2}})^T \\
\mathrm{a}' &= (\mathrm{a}',, \ldots, \alpha'_{\frac{L-1}{2}})^T.
\end{aligned}$$

For notational convenience and to correspond to Ref. [1, Section 11], define

$$L_{\rho,\mathrm{uc}} = \nabla^2 D_{\rho,\mathrm{uc}}(\bar{\rho}) \tag{49}$$

$$\text{and} \quad L_{\rho,\mathrm{fd}} = \nabla^2 D_{\rho,\mathrm{fd}}(\bar{\rho}). \tag{50}$$

Now compute the third partial derivatives of D_ρ. First compute the third partial derivatives of η_ρ, then the third partial derivatives of D_ρ in terms of the derivatives of η_ρ, and

finally an expression for the third partial derivative of D_ρ in terms of H_A and C. Then specialize to C_{uc} and C_{fd}.

Taking a derivative of Eq. **36** gives

$$
\begin{aligned}
(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_\rho) \;=\; & -3(\partial_{r_0}C)^{-5}(\partial_{r_0}\partial_{r_0}C)^2(\partial_{r_i}C)(\partial_{r_j}C)(\partial_{r_k}C) \\
& + (\partial_{r_0}C)^{-4}(\partial_{r_0}\partial_{r_0}\partial_{r_0}C)(\partial_{r_i}C)(\partial_{r_j}C)(\partial_{r_k}C) \\
& + (\partial_{r_0}C)^{-4}(\partial_{r_0}\partial_{r_0}C)\Big[3(\partial_{r_i}\partial_{r_0}C)(\partial_{r_j}C)(\partial_{r_k}C) \\
& \quad + \left[(\partial_{r_0}\partial_{r_j}C) + 2(\partial_{r_j}\partial_{r_0}C)\right](\partial_{r_i}C)(\partial_{r_k}C) + 3(\partial_{r_0}\partial_{r_k}C)(\partial_{r_i}C)(\partial_{r_j}C)\Big] \\
& - (\partial_{r_0}C)^{-3}\Big[(\partial_{r_i}\partial_{r_0}\partial_{r_0}C)(\partial_{r_j}C)(\partial_{r_k}C) + (\partial_{r_0}\partial_{r_j}\partial_{r_0}C)(\partial_{r_i}C)(\partial_{r_k}C) \\
& \quad + (\partial_{r_0}\partial_{r_0}\partial_{r_k}C)(\partial_{r_i}C)(\partial_{r_j}C)\Big] \\
& - (\partial_{r_0}C)^{-3}(\partial_{r_0}\partial_{r_0}C)\left[(\partial_{r_i}\partial_{r_j}C)(\partial_{r_k}C) + (\partial_{r_i}\partial_{r_k}C)(\partial_{r_j}C) + (\partial_{r_j}\partial_{r_k}C)(\partial_{r_i}C)\right] \\
& - (\partial_{r_0}C)^{-3}\Big[2(\partial_{r_i}\partial_{r_0}C)(\partial_{r_j}\partial_{r_0}C)(\partial_{r_k}C) + 2(\partial_{r_i}\partial_{r_0}C)(\partial_{r_0}\partial_{r_k}C)(\partial_{r_j}C) \\
& \quad + \left[(\partial_{r_j}\partial_{r_0}C)(\partial_{r_0}\partial_{r_k}C) + (\partial_{r_0}\partial_{r_j}C)(\partial_{r_0}\partial_{r_k}C)\right](\partial_{r_i}C)\Big] \\
& + (\partial_{r_0}C)^{-2}\left[(\partial_{r_i}\partial_{r_j}\partial_{r_0}C)(\partial_{r_k}C) + (\partial_{r_i}\partial_{r_0}\partial_{r_k}C)(\partial_{r_j}C) + (\partial_{r_0}\partial_{r_j}\partial_{r_k}C)(\partial_{r_i}C)\right] \\
& + (\partial_{r_0}C)^{-2}\left[(\partial_{r_j}\partial_{r_k}C)(\partial_{r_i}\partial_{r_0}C) + (\partial_{r_i}\partial_{r_k}C)(\partial_{r_j}\partial_{r_0}C) + (\partial_{r_i}\partial_{r_j}C)(\partial_{r_0}\partial_{r_k}C)\right] \\
& - (\partial_{r_0}C)^{-1}(\partial_{r_i}\partial_{r_j}\partial_{r_k}C). \qquad (51)
\end{aligned}
$$

As before, if the partial derivatives of C do not depend on the order in which they are taken (i.e., $(\partial_{r_i}\partial_{r_j}C) = (\partial_{r_j}\partial_{r_i}C)$ and so forth) then the partial derivatives of η_ρ are also independent of the order.

Take a derivative of Eq. **38** to get

$$
\begin{aligned}
(\partial_{r_i}\partial_{r_j}\partial_{r_k}D_\rho) \;=\; & (\partial_{r_0}\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\eta_\rho)(\partial_{r_j}\eta_\rho)(\partial_{r_k}\eta_\rho) \\
& + (\partial_{r_i}\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\partial_{r_j}\eta_\rho)(\partial_{r_k}\eta_\rho) + (\partial_{r_0}\partial_{r_j}\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\eta_\rho)(\partial_{r_k}\eta_\rho) \\
& \quad + (\partial_{r_0}\partial_{r_0}\partial_{r_k}\beta H_\lambda)(\partial_{r_i}\eta_\rho)(\partial_{r_j}\eta_\rho) \\
& + (\partial_{r_i}\partial_{r_j}\partial_{r_0}\beta H_\lambda)(\partial_{r_k}\eta_\rho) + (\partial_{r_0}\partial_{r_j}\partial_{r_k}\beta H_\lambda)(\partial_{r_i}\eta_\rho) + (\partial_{r_i}\partial_{r_0}\partial_{r_k}\beta H_\lambda)(\partial_{r_j}\eta_\rho)
\end{aligned}
$$

$$
\begin{aligned}
&+ (\partial_{r_i}\partial_{r_j}\partial_{r_k}\beta H_\lambda) \\
&+ (\partial_{r_0}\partial_{r_0}\beta H_\lambda)\left[(\partial_{r_i}\partial_{r_j}\eta_\rho)(\partial_{r_k}\eta_\rho) + (\partial_{r_i}\partial_{r_k}\eta_\rho)(\partial_{r_j}\eta_\rho) + (\partial_{r_j}\partial_{r_k}\eta_\rho)(\partial_{r_i}\eta_\rho)\right] \\
&+ (\partial_{r_j}\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\partial_{r_k}\eta_\rho) + (\partial_{r_i}\partial_{r_0}\beta H_\lambda)(\partial_{r_j}\partial_{r_k}\eta_\rho) + (\partial_{r_0}\partial_{r_k}\beta H_\lambda)(\partial_{r_i}\partial_{r_j}\eta_\rho) \\
&+ (\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_\rho). \qquad (52)
\end{aligned}
$$

For the problems of interest in this paper, all third partial derivatives of C are zero (**C** is a quadratic form) and all mixed partial derivatives of H_λ are zero. Specializing Eqs. **51** and **52** to this case gives the results that

$$
\begin{aligned}
(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_\rho) \;=\; & -3(\partial_{r_0}C)^{-5}(\partial_{r_0}\partial_{r_0}C)^2(\partial_{r_i}C)(\partial_{r_j}C)(\partial_{r_k}C) \\
& + (\partial_{r_0}C)^{-4}(\partial_{r_0}\partial_{r_0}C)\Big[3(\partial_{r_i}\partial_{r_0}C)(\partial_{r_j}C)(\partial_{r_k}C) \\
& \quad + \left[(\partial_{r_0}\partial_{r_j}C) + 2(\partial_{r_j}\partial_{r_0}C)\right](\partial_{r_i}C)(\partial_{r_k}C) + 3(\partial_{r_0}\partial_{r_k}C)(\partial_{r_i}C)(\partial_{r_j}C)\Big] \\
& - (\partial_{r_0}C)^{-3}(\partial_{r_0}\partial_{r_0}C)\left[(\partial_{r_i}\partial_{r_j}C)(\partial_{r_k}C) + (\partial_{r_i}\partial_{r_k}C)(\partial_{r_j}C) + (\partial_{r_j}\partial_{r_k}C)(\partial_{r_i}C)\right] \\
& - (\partial_{r_0}C)^{-3}\Big[2(\partial_{r_i}\partial_{r_0}C)(\partial_{r_j}\partial_{r_0}C)(\partial_{r_k}C) + 2(\partial_{r_i}\partial_{r_0}C)(\partial_{r_0}\partial_{r_k}C)(\partial_{r_j}C) \\
& \quad + \left[(\partial_{r_j}\partial_{r_0}C)(\partial_{r_0}\partial_{r_k}C) + (\partial_{r_0}\partial_{r_j}C)(\partial_{r_0}\partial_{r_k}C)\right](\partial_{r_i}C)\Big] \\
& + (\partial_{r_0}C)^{-2}\left[(\partial_{r_j}\partial_{r_k}C)(\partial_{r_i}\partial_{r_0}C) + (\partial_{r_i}\partial_{r_k}C)(\partial_{r_j}\partial_{r_0}C) + (\partial_{r_i}\partial_{r_j}C)(\partial_{r_0}\partial_{r_k}C)\right] \qquad (53)
\end{aligned}
$$

and

$$
\begin{aligned}
(\partial_{r_i}\partial_{r_j}\partial_{r_k}D_\rho) \;=\; & (\partial_{r_0}\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\eta_\rho)(\partial_{r_j}\eta_\rho)(\partial_{r_k}\eta_\rho) \\
& + \delta_{i,j}\delta_{j,k}(\partial_{r_i}\partial_{r_i}\partial_{r_i}\beta H_\lambda) \\
& + (\partial_{r_0}\partial_{r_0}\beta H_\lambda)\left[(\partial_{r_i}\partial_{r_j}\eta_\rho)(\partial_{r_k}\eta_\rho) + (\partial_{r_i}\partial_{r_k}\eta_\rho)(\partial_{r_j}\eta_\rho) + (\partial_{r_j}\partial_{r_k}\eta_\rho)(\partial_{r_i}\eta_\rho)\right] \\
& + (\partial_{r_0}\beta H_\lambda)(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_\rho). \qquad (54)
\end{aligned}
$$

Now specialize Eqs. **35,** 36, and **53** to the case where the mixed second partial derivatives of C are zero. Because this case includes C_{uc} but excludes C_{fd}, change notation from "C" to "C_{uc}". The results are

$$
(\partial_{r_i}\eta_{\rho,uc}) \;=\; -(\partial_{r_0}C_{uc})^{-1}(\partial_{r_i}C_{uc}) \qquad (55)
$$

$$(\partial_{r_i}\partial_{r_j}\eta_{\rho,\mathrm{uc}}) = -(\partial_{r_0}C_{\mathrm{uc}})^{-3}(\partial_{r_0}\partial_{r_0}C_{\mathrm{uc}})(\partial_{r_i}C_{\mathrm{uc}})(\partial_{r_j}C_{\mathrm{uc}}) - \delta_{i,j}(\partial_{r_0}C_{\mathrm{uc}})^{-1}(\partial_{r_i}\partial_{r_i}C_{\mathrm{uc}}) \tag{56}$$

$$(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_{\rho,\mathrm{uc}}) = -3(\partial_{r_0}C_{\mathrm{uc}})^{-5}(\partial_{r_0}\partial_{r_0}C_{\mathrm{uc}})^2(\partial_{r_i}C_{\mathrm{uc}})(\partial_{r_j}C_{\mathrm{uc}})(\partial_{r_k}C_{\mathrm{uc}}) - (\partial_{r_0}C_{\mathrm{uc}})^{-3}(\partial_{r_0}\partial_{r_0}C_{\mathrm{uc}})\Big[\delta_{i,j}(\partial_{r_i}\partial_{r_i}C_{\mathrm{uc}})(\partial_{r_k}C_{\mathrm{uc}}) + \delta_{i,k}(\partial_{r_i}\partial_{r_i}C_{\mathrm{uc}})(\partial_{r_j}C_{\mathrm{uc}}) + \delta_{j,k}(\partial_{r_j}\partial_{r_j}C_{\mathrm{uc}})(\partial_{r_i}C_{\mathrm{uc}})\Big]. \tag{57}$$

Substituting the actud values for C_{uc} and its derivatives d l evaluated at the critical point $\bar{r} = \bar{\rho}$ gives the results:

$$(\partial_{r_i}\eta_{\rho,\mathrm{uc}}) = -\frac{\frac{4}{L}\rho_i}{\frac{2}{L}\rho_0 - 1} \tag{58}$$

$$(\partial_{r_i}\partial_{r_j}\eta_{\rho,\mathrm{uc}}) = -\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2 \rho_i\rho_j - \delta_{i,j}\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1} \tag{59}$$

$$(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_{\rho,\mathrm{uc}}) = -3\left(\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^3 \rho_i\rho_j\rho_k - \frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2 [\delta_{i,j}\rho_k + \delta_{i,k}\rho_j + \delta_{j,k}\rho_i] \tag{60}$$

$$(\partial_{r_i}\partial_{r_j}\eta_{\rho,\mathrm{uc}})(\partial_{r_k}\eta_{\rho,\mathrm{uc}}) + (\partial_{r_i}\partial_{r_k}\eta_{\rho,\mathrm{uc}})(\partial_{r_j}\eta_{\rho,\mathrm{uc}}) + (\partial_{r_j}\partial_{r_k}\eta_{\rho,\mathrm{uc}})(\partial_{r_i}\eta_{\rho,\mathrm{uc}}) = 3\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^3 \rho_i\rho_j\rho_k + \left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2 (\delta_{i,j}\rho_k + \delta_{i,k}\rho_j + \delta_{j,k}\rho_i)$$

and

$$\left(\partial_{r_i}\partial_{r_j}\partial_{r_k}D_{\rho,\mathrm{uc}}\right) = \kappa_3\rho_i\rho_j\rho_k + \kappa_2[\rho_i\delta_{j,k} + \rho_j\delta_{i,k} + \rho_k\delta_{i,j}] + \kappa_1\delta_{i,j}\delta_{j,k} \tag{61}$$

where

$$\kappa_3 = \left[-(\partial_{r_0}\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\rho_0) + (\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\rho_0)\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}3 - (\partial_{r_0}\beta H_\lambda)(\rho_0)\left(\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2 3\right]\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^3$$

$$\kappa_2 = \left[(\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\rho_0) - (\partial_{r_0}\beta H_\lambda)(\rho_0)\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1}\right]\left(\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1}\right)^2$$

$$\kappa_1 = (\partial_{r_k}\partial_{r_k}\partial_{r_k}\beta H_\lambda)(\rho_k).$$

Return to Eq. 54 and consider the case where $C = C_{\rm fd}$. Eq. 35 is used unchanged except for replacing "C" by "$C_{\rm fd}$". Eq. 36 specializes to

$$\begin{aligned}(\partial_{r_i}\partial_{r_j}\eta_{\rho,\rm fd})(\bar\rho) &= -(\partial_{r_0}C_{\rm fd})^{-3}\frac{1}{L}\left(1+\frac{1}{L}\right)(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})\\ &\quad+(\partial_{r_0}C_{\rm fd})^{-2}\frac{2}{L^2}\left[(\partial_{r_j}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})\right]\\ &\quad-(\partial_{r_0}C_{\rm fd})^{-1}\left(\frac{4}{L^2}+\delta_{i,j}\frac{2}{L}\right).\end{aligned}\tag{62}$$

Combining Eqs. 35 and 62 gives

$$\begin{aligned}&(\partial_{r_i}\partial_{r_j}\eta_{\rho,\rm fd})(\bar\rho)(\partial_{r_k}\eta_{\rho,\rm fd})(\bar\rho)+(\partial_{r_i}\partial_{r_k}\eta_{\rho,\rm fd})(\bar\rho)(\partial_{r_j}\eta_{\rho,\rm fd})(\bar\rho)+(\partial_{r_j}\partial_{r_k}\eta_{\rho,\rm fd})(\bar\rho)(\partial_{r_i}\eta_{\rho,\rm fd})(\bar\rho)\\ &\quad=(\partial_{r_0}C_{\rm fd})^{-4}\frac{1}{L}\left(1+\frac{1}{L}\right)3(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})(\partial_{r_k}C_{\rm fd})\\ &\qquad-(\partial_{r_0}C_{\rm fd})^{-3}\frac{2}{L^2}2\left[(\partial_{r_j}C_{\rm fd})(\partial_{r_k}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})(\partial_{r_k}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})\right]\\ &\qquad+(\partial_{r_0}C_{\rm fd})^{-2}\frac{4}{L^2}\left[(\partial_{r_k}C_{\rm fd})+(\partial_{r_j}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})\right]\\ &\qquad+(\partial_{r_0}C_{\rm fd})^{-2}\frac{2}{L}\left[\delta_{i,j}(\partial_{r_k}C_{\rm fd})+\delta_{i,k}(\partial_{r_j}C_{\rm fd})+\delta_{j,k}(\partial_{r_i}C_{\rm fd})\right]\end{aligned}\tag{63}$$

and

$$(\partial_{r_i}\eta_{\rho,\rm fd})(\bar\rho)(\partial_{r_j}\eta_{\rho,\rm fd})(\bar\rho)(\partial_{r_k}\eta_{\rho,\rm fd})(\bar\rho) = -(\partial_{r_0}C_{\rm fd})^{-3}(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})(\partial_{r_k}C_{\rm fd}).\tag{64}$$

Finally, specializing Eq. 53 gives

$$\begin{aligned}&(\partial_{r_i}\partial_{r_j}\partial_{r_k}\eta_{\rho,\rm fd})(\bar\rho)\\ &\quad=-3(\partial_{r_0}C_{\rm fd})^{-5}\left[\frac{1}{L}\left(1+\frac{1}{L}\right)\right]^2(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})(\partial_{r_k}C_{\rm fd})\\ &\qquad+(\partial_{r_0}C_{\rm fd})^{-4}\frac{1}{L}\left(1+\frac{1}{L}\right)3\frac{2}{L^2}\left[(\partial_{r_j}C_{\rm fd})(\partial_{r_k}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})(\partial_{r_k}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})(\partial_{r_j}C_{\rm fd})\right]\\ &\qquad-(\partial_{r_0}C_{\rm fd})^{-3}\left[\frac{1}{L}\left(1+\frac{1}{L}\right)\frac{4}{L^2}+2\left(\frac{2}{L^2}\right)^2\right]\left[(\partial_{r_k}C_{\rm fd})+(\partial_{r_j}C_{\rm fd})+(\partial_{r_i}C_{\rm fd})\right]\\ &\qquad-(\partial_{r_0}C_{\rm fd})^{-3}\frac{1}{L}\left(1+\frac{1}{L}\right)\frac{2}{L}\left[\delta_{i,j}(\partial_{r_k}C_{\rm fd})+\delta_{i,k}(\partial_{r_j}C_{\rm fd})+\delta_{j,k}(\partial_{r_i}C_{\rm fd})\right]\\ &\qquad+(\partial_{r_0}C_{\rm fd})^{-2}\frac{2}{L^2}\frac{4}{L^2}3\\ &\qquad+(\partial_{r_0}C_{\rm fd})^{-2}\frac{2}{L^2}\frac{2}{L}\left[\delta_{j,k}+\delta_{i,k}+\delta_{i,j}\right].\end{aligned}\tag{65}$$

Use Eqs. 35, 62, 63, 64, and 65 in Eq. 54 to get

$$
\begin{aligned}
&(\partial_{r_i}\partial_{r_j}\partial_{r_k}D_{\rho,\mathrm{fd}})(\bar{\rho}) \\
&\quad = (\partial_{r_i}C_{\mathrm{fd}})(\partial_{r_j}C_{\mathrm{fd}})(\partial_{r_k}C_{\mathrm{fd}})\kappa'_3 \\
&\qquad + \left[(\partial_{r_j}C_{\mathrm{fd}})(\partial_{r_k}C_{\mathrm{fd}}) + (\partial_{r_i}C_{\mathrm{fd}})(\partial_{r_k}C_{\mathrm{fd}}) + (\partial_{r_i}C_{\mathrm{fd}})(\partial_{r_j}C_{\mathrm{fd}})\right]\kappa'_{2a} \\
&\qquad + \left[(\partial_{r_k}C_{\mathrm{fd}}) + (\partial_{r_j}C_{\mathrm{fd}}) + (\partial_{r_i}C_{\mathrm{fd}})\right]\kappa'_{1a} \\
&\qquad + \left[\delta_{i,j}(\partial_{r_k}C_{\mathrm{fd}}) + \delta_{i,k}(\partial_{r_j}C_{\mathrm{fd}}) + \delta_{j,k}(\partial_{r_i}C_{\mathrm{fd}})\right]\kappa'_{2b} \\
&\qquad + \left[\delta_{j,k} + \delta_{i,k} + \delta_{i,j}\right]\kappa'_{1b} \\
&\qquad + \kappa'_0 \\
&\qquad + \delta_{i,j}\delta_{j,k}\kappa'_{2c}
\end{aligned}
$$

where

$$
\begin{aligned}
\kappa'_3 &= (\partial_{r_0}C_{\mathrm{fd}})^{-3}\Big\{-(\partial_{r_0}\partial_{r_0}\partial_{r_0}\beta H_\lambda) + (\partial_{r_0}\partial_{r_0}\beta H_\lambda)(\partial_{r_0}C_{\mathrm{fd}})^{-1}\frac{1}{L}\left(1+\frac{1}{L}\right)3 \\
&\qquad - (\partial_{r_0}\beta H_\lambda)3(\partial_{r_0}C_{\mathrm{fd}})^{-2}\left[\frac{1}{L}\left(1+\frac{1}{L}\right)\right]^2\Big\} \\
\kappa'_{2a} &= (\partial_{r_0}C_{\mathrm{fd}})^{-3}\frac{2}{L^2}\left[-2(\partial_{r_0}\partial_{r_0}\beta H_\lambda) + (\partial_{r_0}\beta H_\lambda)(\partial_{r_0}C_{\mathrm{fd}})^{-1}\frac{1}{L}\left(1+\frac{1}{L}\right)3\right] \\
\kappa'_{1a} &= (\partial_{r_0}C_{\mathrm{fd}})^{-2}\frac{4}{L^2}\left\{(\partial_{r_0}\partial_{r_0}\beta H_\lambda) - (\partial_{r_0}\beta H_\lambda)(\partial_{r_0}C_{\mathrm{fd}})^{-1}\left[\frac{1}{L}\left(1+\frac{1}{L}\right)+\frac{2}{L^2}\right]\right\} \\
\kappa'_{2b} &= (\partial_{r_0}C_{\mathrm{fd}})^{-2}\frac{2}{L}\left[(\partial_{r_0}\partial_{r_0}\beta H_\lambda) - (\partial_{r_0}\beta H_\lambda)(\partial_{r_0}C_{\mathrm{fd}})^{-1}\frac{1}{L}\left(1+\frac{1}{L}\right)\right] \\
\kappa'_{1b} &= (\partial_{r_0}C_{\mathrm{fd}})^{-2}\frac{2}{L^2}\frac{2}{L}(\partial_{r_0}\beta H_\lambda) \\
\kappa'_0 &= (\partial_{r_0}C_{\mathrm{fd}})^{-2}\frac{2}{L^2}\frac{4}{L^2}3(\partial_{r_0}\beta H_\lambda) \\
\kappa'_{2c} &= (\partial_{r_i}\partial_{r_i}\partial_{r_i}\beta H_\lambda).
\end{aligned}
$$

23 Appendix–Derivatives of $g_k(F_\rho(\bar{r}))$ for Approaches 2 and 3

In this appendix the first and second mixed partial derivatives of $g_k(F_\rho(\bar{r}))$ with respect to r_l for $l > 0$ for both C_{uc} and C_{fd} are computed.

Using Eqs. 20 and 19 in Eqs. 15 and 16 gives

$$g_Z(F_\rho(\bar{r})) = \exp(-\beta h_{0,0}(\eta_\rho(\bar{r}))) \exp\left(\sum_{l=1}^{\frac{L-1}{2}} -\beta h_{l,0}(r_l)\right)$$

$$g_k(F_\rho(\bar{r})) = \begin{cases} \eta_\rho(\bar{r}) g_Z(F_\rho(\bar{r})), & k = 0 \\ r_k g_Z(F_\rho(\bar{r})), & k \neq 0 \end{cases}$$

$$= \begin{cases} \eta_\rho(\bar{r}) \exp(-\beta h_{0,0}(\eta_\rho(\bar{r}))) \exp\left(\sum_{l=1}^{\frac{L-1}{2}} -\beta h_{l,0}(r_l)\right), & k = 0 \\ r_k \exp(-\beta h_{0,0}(\eta_\rho(\bar{r}))) \exp\left(\sum_{l=1}^{\frac{L-1}{2}} -\beta h_{l,0}(r_l)\right), & k \neq 0 \end{cases}$$

Taking derivatives with respect to r_j, $j \in K_L^+$ one finds that

$$\partial_{r_j} [g_Z(F_\rho(\bar{r}))] = g_Z(F_\rho(\bar{r}))\omega_j(\bar{r})$$

where

$$\omega_j(\bar{r}) = \beta h'_{0,0}(\eta_\rho(\bar{r}))[(\partial_{r_0} C)(F_\rho(\bar{r}))]^{-1}(\partial_{r_j} C)(F_\rho(\bar{r})) - \beta h'_{j,0}(r_j).$$

Note that $\omega_j(\bar{r})$ depends on Ψ_0 and Ψ_j through $\beta h'_{0,0}$ and $\beta h'_{j,0}$. It is useful to have the formula

$$\begin{aligned}(\partial_{r_i}\omega_j)(\bar{r}) = {} & \left[\beta h'_{0,0}(\eta_\rho(\bar{r}))(\partial_{r_0}\partial_{r_0} C)(F_\rho(\bar{r}))[(\partial_{r_0} C)(F_\rho(\bar{r}))]^{-1} - \beta h''_{0,0}(\eta_\rho(\bar{r}))\right] \times \\ & \times [(\partial_{r_0} C)(F_\rho(\bar{r}))]^{-2}(\partial_{r_i} C)(F_\rho(\bar{r}))(\partial_{r_j} C)(F_\rho(\bar{r})) \\ & - \beta h'_{0,0}(\eta_\rho(\bar{r}))[(\partial_{r_0} C)(F_\rho(\bar{r}))]^{-2} \times \\ & \times \left[(\partial_{r_i}\partial_{r_0} C)(F_\rho(\bar{r}))(\partial_{r_j} C)(F_\rho(\bar{r})) + (\partial_{r_0}\partial_{r_j} C)(F_\rho(\bar{r}))(\partial_{r_i} C)(F_\rho(\bar{r}))\right] \\ & + \beta h'_{0,0}(\eta_\rho(\bar{r}))[(\partial_{r_0} C)(F_\rho(\bar{r}))]^{-1}(\partial_{r_i}\partial_{r_j} C)(F_\rho(\bar{r})) \\ & - \delta_{i,j}\beta h''_{j,0}(r_j).\end{aligned}$$

Note that $(\partial_{r_i}\omega_j)(\bar{r})$ depends on Ψ_0 through $\beta h'_{0,0}$ but does not depend on Ψ_j since $\beta h''_{j,0}$ is independent of Ψ_j. Note that $(\partial_{r_i}\omega_j)(\bar{r}) = (\partial_{r_j}\omega_i)(\bar{r})$.

Considering only $k \neq 0$, compute the first and second derivatives of $g_k(F_\rho(\bar{r}))$:

$$\partial_{r_j} [g_k(F_\rho(\bar{r}))] = \delta_{j,k} g_Z(F_\rho(\bar{r})) + r_k g_Z(F_\rho(\bar{r}))\omega_j(\bar{r})$$

$$\partial_{r_i}\partial_{r_j} [g_k(F_\rho(\bar{r}))] = [\delta_{i,k}\omega_j(\bar{r}) + \delta_{j,k}\omega_i(\bar{r})] g_Z(F_\rho(\bar{r})) + r_k g_Z(F_\rho(\bar{r}))[\omega_i(\bar{r})\omega_j(\bar{r}) + (\partial_{r_i}\omega_j)(\bar{r})].$$

Note that $\partial_{r_j}[g_k(F_\rho(\bar{r}))]$ and $\partial_{r_i}\partial_{r_j}[g_k(F_\rho(\bar{r}))]$ both depend on Ψ_k for all k. Note that

$$\partial_{r_i}\partial_{r_j}[g_k(F_\rho(\bar{r}))] = \partial_{r_j}\partial_{r_i}[g_k(F_\rho(\bar{r}))]$$

because $(\partial_{r_i}\omega_j)(\bar{r}) = (\partial_{r_j}\omega_i)(\bar{r})$.

Evaluate these results for $\bar{r} = \bar{\rho}$:

$$\begin{aligned}
g_Z(F_\rho(\bar{\rho})) &= g_Z(\rho) \\
&= \exp\left(\sum_{l=0}^{\frac{L-1}{2}} -\beta h_{l,0}(\rho_l)\right) \\
\omega_j(\bar{\rho}) &= \beta h'_{0,0}(\rho_0)[(\partial_{r_0}C)(\rho)]^{-1}(\partial_{r_j}C)(\rho) - \beta h'_{j,0}(\rho_j) \\
(\partial_{r_i}\omega_j)(\bar{\rho}) &= \left[\beta h'_{0,0}(\rho_0)(\partial_{r_0}\partial_{r_0}C)(\rho)[(\partial_{r_0}C)(\rho)]^{-1} - \beta h''_{0,0}(\rho_0)\right] \times \\
&\qquad \times [(\partial_{r_0}C)(\rho)]^{-2}(\partial_{r_i}C)(\rho)(\partial_{r_j}C)(\rho) \\
&\quad - \beta h'_{0,0}(\rho_0)[(\partial_{r_0}C)(\rho)]^{-2} \times \\
&\qquad \times \left[(\partial_{r_i}\partial_{r_0}C)(\rho)(\partial_{r_j}C)(\rho) + (\partial_{r_0}\partial_{r_j}C)(\rho)(\partial_{r_i}C)(\rho)\right] \\
&\quad + \beta h'_{0,0}(\rho_0)[(\partial_{r_0}C)(\rho)]^{-1}(\partial_{r_i}\partial_{r_j}C)(\rho) \\
&\quad - \delta_{i,j}\beta h''_{j,0}(\rho_j) \\
\partial_{r_j}[g_k(F_\rho(\bar{r}))]\Big|_{\bar{\rho}} &= \delta_{j,k}g_Z(\rho) + \rho_k g_Z(\rho)\omega_j(\bar{\rho}) \\
\partial_{r_i}\partial_{r_j}[g_k(F_\rho(\bar{r}))]\Big|_{\bar{\rho}} &= [\delta_{i,k}\omega_j(\bar{\rho}) + \delta_{j,k}\omega_i(\bar{\rho})]\, g_Z(\rho) + \rho_k g_Z(\rho)\,[\omega_i(\bar{\rho})\omega_j(\bar{\rho}) + (\partial_{r_i}\omega_j)(\bar{\rho})]\,.
\end{aligned}$$

Note that only $\omega_i(\bar{\rho})$ and $(\partial_{r_j}\omega_i)(\bar{\rho})$ have explicit dependence on C. The other functions depend on C only indirectly throught ρ, $\omega_i(\bar{\rho})$, and $(\partial_{r_j}\omega_i)(\bar{\rho})$.

Evaluate $\omega_i(\bar{\rho})$ and $(\partial_{r_j}\omega_i)(\bar{\rho})$ for C_{uc} and C_{fd}. For the case of C_{uc}:

$$\begin{aligned}
\omega_{j,\text{uc}}(\rho_0,\rho_j) &= \beta h'_{0,0}(\rho_0)\frac{\frac{4}{L}\rho_j}{\frac{2}{L}\rho_0 - 1} - \beta h'_{j,0}(\rho_j) \qquad (66) \\
(\partial_{r_i}\omega_{j,\text{uc}})(\rho_0,\rho_i,\rho_j) &= \left[\beta h'_{0,0}(\rho_0)\frac{\frac{2}{L}}{\frac{2}{L}\rho_0 - 1} - \beta h''_{0,0}(\rho_0)\right]\frac{\frac{4}{L}\rho_i\frac{4}{L}\rho_j}{\left(\frac{2}{L}\rho_0 - 1\right)^2} \\
&\quad + \delta_{i,j}\left[\beta h'_{0,0}(\rho_0)\frac{\frac{4}{L}}{\frac{2}{L}\rho_0 - 1} - \beta h''_{j,0}(\rho_j)\right].
\end{aligned}$$

For the case of C_{fd}:

$$\begin{aligned}
\omega_{j,\mathrm{fd}}(\bar{\rho}) &= \beta h'_{0,0}(\rho_0)\frac{2(2\rho_j + 4v^T\rho - 1)}{2\rho_0 + 4v^T\rho - 1 - L} - \beta h'_{j,0}(\rho_j) \\
(\partial_{r_i}\omega_{j,\mathrm{fd}})(\bar{\rho}) &= \left[\beta h'_{0,0}(\rho_0)\frac{2\left(1+\frac{1}{L}\right)}{2\rho_0 + 4v^T\rho - 1 - L} - \beta h''_{0,0}(\rho_0)\right] \times \\
&\quad \times \frac{4(2\rho_i + 4v^T\rho - 1)(2\rho_j + 4v^T\rho - 1)}{(2\rho_0 + 4v^T\rho - 1 - L)^2} \\
&\quad - \beta h'_{0,0}(\rho_0)\frac{16(\rho_i + \rho_j + 4v^T\rho - 1)}{L\,(2\rho_0 + 4v^T\rho - 1 - L)^2} \\
&\quad + \beta h'_{0,0}(\rho_0)\frac{\frac{8}{L} + \delta_{i,j}4}{2\rho_0 + 4v^T\rho - 1 - L} \\
&\quad - \delta_{i,j}\beta h''_{j,0}(\rho_j).
\end{aligned}$$

Specialize to the case where $\rho_k = 0$. In general the expressions for $\omega_{i,\mathrm{uc}}(\rho_0, \rho_i)$, $(\partial_{r_j}\omega_{i,\mathrm{uc}})(\rho_0, \mathrm{pi}, \rho_j)$, $\omega_{i,\mathrm{fd}}(\rho_0, \rho_i)$, and $(\partial_{r_j}\omega_{i,\mathrm{fd}})(\rho_0, \rho_i, \rho_j)$ do not simplify (except in the special cases where $i = \mathrm{k}$, $\mathrm{j} = \mathrm{k}$, or $i = \mathrm{j} = \mathrm{k}$). Similarly, the expression for $g_Z(\rho)$ does not have significant changes. However, for any C, $\partial_{r_j}\left[g_k(F_\rho(\bar{r}))\right]\Big|_{\mathrm{p}}$ and $\partial_{r_i}\partial_{r_j}\left[g_k(F_\rho(\bar{r}))\right]\Big|_{\rho}$ are greatly simpified. Specifically,

$$\partial_{r_j}\left[g_k(F_\rho(\bar{r}))\right]\Big|_{\rho} = \delta_{j,k}g_Z(\rho) \tag{67}$$

$$\partial_{r_i}\partial_{r_j}\left[g_k(F_\rho(\bar{r}))\right]\Big|_{\rho} = \left[\delta_{i,k}\omega_j(\bar{\rho}) + \delta_{j,k}\omega_i(\bar{\rho})\right] g_Z(\rho). \tag{68}$$

In the case where $\rho_k = 0$, define the vector $q_{\rho,k}$ and the matrix $Q_{\rho,k}$ by

$$(q_{\rho,k})_j = \delta_{j,k}g_Z(\rho) \tag{69}$$

$$(Q_{\rho,k})_{i,j} = \left[\delta_{i,k}\omega_j(\bar{\rho}) + \delta_{j,k}\omega_i(\bar{\rho})\right] g_Z(\rho). \tag{70}$$

24 Appendix–$J_{1,c}$ Formulae for Approaches 2 and 3

In this appendix $J_{1,c}$ is computed for Approaches 2 and 3. Because $J_{1,c}$ is a contribution to $E(\Phi_k|y)$, there is a k dependence for $J_{1,c}$ that is suppressed in the notation. As is seen in the following, for Approach 2 $J_{1,c}$ is zero for any k while for Approach 3 it in typically nonzero

for all k. The term $J_{1,c}$ determines a part of the second order contribution, which is required only when $\rho_k = 0$ so only that case is considered.

The quantity $J_{1,c}$ is defined in Eq. 82 which, in terms of the standard notation, is

$$J_{1,c} = -\frac{1}{2}\sum_{i,j,l,m}\left(\partial_{r_i}\partial_{r_j}\partial_{r_l}D_\rho\right)(\rho)\left(\partial_{r_m}[g_k(F_\rho(\bar{r}))]\right)|_\rho\left(L_\rho^{-1}\right)_{i,j}\left(L_\rho^{-1}\right)_{l,m}. \tag{71}$$

Since $\rho_k = 0$, Eq. 67 applies which leads to the simplified form

$$J_{1,c} = -\frac{g_Z(\rho)}{2}\sum_{i,j,l}\left(\partial_{r_i}\partial_{r_j}\partial_{r_l}D_\rho\right)(\rho)\left(L_\rho^{-1}\right)_{i,j}\left(L_\rho^{-1}\right)_{l,k}. \tag{72}$$

Therefore, $J_{1,c}$ is proportional to $g_Z(\rho)$ and depends on Ψ only through $g_Z(\rho)$.

Consider Approach 2. $\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,j}$ is defined in Eqs. 49 and 45 which in turn depend on Eqs. 44, 21, 46 and 47. Note especially that the second term of Eq. 45 is proportional to $\gamma_{i,\mathrm{uc}}$ (defined in Eq. 46) which in turn is proportional to ρi. Therefore, since $\rho_k = 0$,

$$\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{l,k} = \delta_{l,k}\frac{1}{e_{l,\mathrm{uc}}}.$$

Using this result in Eq. 72 gives

$$J_{1,c,\mathrm{uc}} = -\frac{g_Z(\rho)}{2e_{k,\mathrm{uc}}}\sum_{i,j}\left(\partial_{r_i}\partial_{r_j}\partial_{r_k}D_{\rho,\mathrm{uc}}\right)(\rho)\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,j}.$$

Because $\rho_k = 0$, the $\kappa_3\rho_i\rho_j\rho_k$ (because $\rho_k = 0$) and $\kappa_1\delta_{i,j}\delta_{j,l}$ (because $\kappa_1 = 0$) terms are zero in Eq. 61. Use this result in the previous expression for $J_{1,c,\mathrm{uc}}$ to get

$$\begin{aligned} J_{1,c,\mathrm{uc}} &= -\frac{g_Z(\rho)}{2e_{k,\mathrm{uc}}}\sum_{i,j}\kappa_2\left[\rho_i\delta_{j,k}+\rho_j\delta_{i,k}+\rho_k\delta_{i,j}\right]\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,j} \\ &= -\frac{g_Z(\rho)}{2e_{k,\mathrm{uc}}}\sum_{i,j}\kappa_2\left[\rho_i\delta_{j,k}+\rho_j\delta_{i,k}\right]\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,j} \end{aligned}$$

since $\rho_k = 0$. Simplifying gives

$$J_{1,c,\mathrm{uc}} = -\frac{\kappa_2 g_Z(\rho)}{e_{k,\mathrm{uc}}}\sum_i \rho_i\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,k}$$

since $\left(L_{\rho,\mathrm{uc}}^{-1}\right)_{i,j} = \left(L_{\rho,\mathrm{uc}}^{-1}\right)_{j,i}$. Finally taking the i, k element of Eqs. 49 and 45 with $\gamma_{k,\mathrm{uc}} = \frac{\rho_k}{e_{k,\mathrm{uc}}}$, and $\rho_k = 0$ gives

$$J_{1,c,\mathrm{uc}} = -\frac{\kappa_2 g_Z(\rho)}{e_{k,\mathrm{uc}}}\sum_i \rho_i\left(\delta_{i,k}\frac{1}{e_{i,\mathrm{uc}}} - \frac{1}{\sum_{k'=1}^{\frac{L-1}{2}}\frac{\rho_{k'}^2}{e_{k',\mathrm{uc}}}+R}\gamma_{i,\mathrm{uc}}\gamma_{k,\mathrm{uc}}\right)$$

$$
\begin{aligned}
&= -\frac{\kappa_2 g_Z(\rho)}{e_{k,\mathrm{uc}}} \sum_i \rho_i \delta_{i,k} \frac{1}{e_{i,\mathrm{uc}}} \\
&= -\frac{\kappa_2 g_Z(\rho)}{e_{k,\mathrm{uc}}^2} \rho_k \\
&= 0
\end{aligned} \tag{73}
$$

as claimed. This completes the calculation for Approach 2.

Consider Approach 3. From Eqs. 48 and 50 it follows that

$$
\begin{aligned}
(L_{\rho,\mathrm{fd}}^{-1})_{i,j}(L_{\rho,\mathrm{fd}}^{-1})_{l,k} &= \delta_{i,j}\delta_{l,k}\frac{1}{e_{i,\mathrm{fd}}e_{l,\mathrm{fd}}} \\
&\quad - \delta_{i,j}\frac{1}{\Delta}\left(\frac{\alpha_l\zeta_k}{e_{i,\mathrm{fd}}} + \frac{\alpha_l'\zeta_k'}{e_{i,\mathrm{fd}}}\right) \\
&\quad - \delta_{l,k}\frac{1}{\Delta}\left(\frac{\alpha_i\zeta_j}{e_{l,\mathrm{fd}}} + \frac{\alpha_i'\zeta_j'}{e_{l,\mathrm{fd}}}\right) \\
&\quad + \frac{1}{\Delta^2}\left(\alpha_i\zeta_j\alpha_l\zeta_k + \alpha_i\zeta_j\alpha_l'\zeta_k' + \alpha_i'\zeta_j'\alpha_l\zeta_k + \alpha_i'\zeta_j'\alpha_l'\zeta_k'\right).
\end{aligned} \tag{74}
$$

Define

$$
\begin{aligned}
S_0 &= \sum_i \frac{1}{e_{i,\mathrm{fd}}} \\
S_1 &= \sum_i (\partial_{r_i} C_{\mathrm{fd}}) \frac{1}{e_{i,\mathrm{fd}}} \\
S_2 &= \sum_i (\partial_{r_i} C_{\mathrm{fd}})^2 \frac{1}{e_{i,\mathrm{fd}}} \\
S_3 &= \sum_i (\partial_{r_i} C_{\mathrm{fd}}) \alpha_i \\
S_3' &= \sum_i (\partial_{r_i} C_{\mathrm{fd}}) \alpha_i' \\
S_4 &= \sum_i \alpha_i \\
S_4' &= \sum_i \alpha_i' \\
S_5 &= \sum_i (\partial_{r_i} C_{\mathrm{fd}}) \frac{\alpha_i}{e_{i,\mathrm{fd}}} \\
S_5' &= \sum_i (\partial_{r_i} C_{\mathrm{fd}}) \frac{\alpha_i'}{e_{i,\mathrm{fd}}} \\
S_6 &= \sum_i \frac{\alpha_i}{e_{i,\mathrm{fd}}} \\
S_6' &= \sum_i \frac{\alpha_i'}{e_{i,\mathrm{fd}}}
\end{aligned}
$$

$$
\begin{aligned}
S_7 &= \sum_i (\partial_{r_i} C_{\mathrm{fd}})\zeta_i \\
S_7' &= \sum_i (\partial_{r_i} C_{\mathrm{fd}})\zeta_i' \\
S_8 &= \sum_i \zeta_i \\
S_8' &= \sum_i \zeta_i' \\
S_9 &= \sum_i \alpha_i \zeta_i \\
S_9'^a &= \sum_i \alpha_i' \zeta_i \\
S_9'^b &= \sum_i \alpha_i \zeta_i' \\
S_9'' &= \sum_i \alpha_i' \zeta_i' \\
S_{10} &= \sum_i \alpha_i \alpha_i \\
S_{10}' &= \sum_i \alpha_i \alpha_i' \\
S_{10}'' &= \sum_i \alpha_i' \alpha_i' \\
S_{11} &= \sum_i \alpha_i^2 \zeta_i \\
S_{11}' &= \sum_i \alpha_i \alpha_i' \zeta_i \\
S_{11}'' &= \sum_i \alpha_i \alpha_i' \zeta_i' \\
S_{11}''' &= \sum_i \alpha_i'^2 \zeta_i'.
\end{aligned}
$$

Note that the computation of each of these S variables is linear in the size of the lattice and therefore is practical. Furthermore, the values of the S variables are the same for any value of k so the linear computation need be done only once per reconstruction. Finally, the values of the S variables are independent of Ψ_l for all l so no form of **symmetry** breaking **optimization** can require recalculation of the S variables.

Substitution of Eq. **74** into Eq. **72** leads to four terms denoted $J_{1,c,\mathrm{fd},1}, \ldots, J_{1,c,\mathrm{fd},4}$. These terms, **simplified** through use of the S variables, are

$$
\begin{aligned}
J_{1,c,\mathrm{fd},1} &= -\frac{gz(\rho)}{2} \sum_i (\partial_{r_i} \partial_{r_i} \partial_{r_k} D_\rho) \frac{1}{e_{i,\mathrm{fd}} e_{k,\mathrm{fd}}} \\
&= -\frac{gz(\rho)}{2e_{k,\mathrm{fd}}} \Big\{ S_2 \left[(\partial_{r_k} C_{\mathrm{fd}}) \kappa_3' + \kappa_{2a}' \right]
\end{aligned}
$$

$$
\begin{aligned}
&+ S_1 2\left[(\partial_{r_k} C_{\mathrm{fd}})\kappa'_{2a} + \kappa'_{1a}\right] \\
&+ S_0\left[\kappa'_{1b} + \kappa'_0 + (\partial_{r_k} C_{\mathrm{fd}})\left(\kappa'_{1a} + \kappa'_{2b}\right)\right] \\
&+ \left[2(\partial_{r_k} C_{\mathrm{fd}})\kappa'_{2b} + 2\kappa'_{1b} + \kappa'_{2c}\right]\frac{1}{e_{k,\mathrm{fd}}}\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
J_{1,c,\mathrm{fd},2} &= \frac{gz(\rho)}{2}\sum_{i,l}(\partial_{r_i}\partial_{r_i}\partial_{r_l} D_\rho)\frac{1}{\Delta}\left(\frac{\alpha_l\zeta_k}{e_{i,\mathrm{fd}}} + \frac{\alpha'_l\zeta'_k}{e_{i,\mathrm{fd}}}\right) \\
&= \frac{gz(\rho)}{2\Delta}\Bigg\{S_2\left(S_3\zeta_k + S'_3\zeta'_k\right)\kappa'_3 \\
&\quad + \left[2S_1\left(S_3\zeta_k + S'_3\zeta'_k\right) + S_2\left(S_4\zeta_k + S'_4\zeta'_k\right)\right]\kappa'_{2a} \\
&\quad + S_0\left(S_3\zeta_k + S'_3\zeta'_k\right)\left(\kappa'_{1a} + \kappa'_{2b}\right) \\
&\quad + 2S_1\left(S_4\zeta_k + S'_4\zeta'_k\right)\kappa'_{1a} \\
&\quad + S_0\left(S_4\zeta_k + S'_4\zeta'_k\right)\left(\kappa'_{1b} + \kappa'_0\right) \\
&\quad + 2\left(S_5\zeta_k + S'_5\zeta'_k\right)\kappa'_{2b} + \left(S_6\zeta_k + S'_6\zeta'_k\right)\left(2\kappa'_{1b} + \kappa'_{2c}\right)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
J_{1,c,\mathrm{fd},3} &= \frac{gz(\rho)}{2}\sum_{i,j}(\partial_{r_i}\partial_{r_j}\partial_{r_k} D_\rho)\frac{1}{\Delta}\left(\frac{\alpha_i\zeta_j}{e_{k,\mathrm{fd}}} + \frac{\alpha'_i\zeta'_j}{e_{k,\mathrm{fd}}}\right) \\
&= \frac{gz(\rho)}{2\Delta}\Bigg\{\left(S_3S_7 + S'_3S'_7\right)\frac{(\partial_{r_k} C_{\mathrm{fd}})\kappa'_3}{e_{k,\mathrm{fd}}} \\
&\quad + \left[\left(S_4S_7 + S'_4S'_7\right)(\partial_{r_k} C_{\mathrm{fd}}) + \left(S_3S_8 + S'_3S'_8\right)(\partial_{r_k} C_{\mathrm{fd}}) + \left(S_3S_7 + S'_3S'_7\right)\right]\frac{\kappa'_{2a}}{e_{k,\mathrm{fd}}} \\
&\quad + \left[\left(S_4S_8 + S'_4S'_8\right)(\partial_{r_k} C_{\mathrm{fd}}) + \left(S_4S_7 + S'_4S'_7\right) + \left(S_3S_8 + S'_3S'_8\right)\right]\frac{\kappa'_{1a}}{e_{k,\mathrm{fd}}} \\
&\quad + \left(S_4S_8 + S'_4S'_8\right)\kappa'_0 \\
&\quad + \left(S_4\zeta_k + S'_4\zeta'_k\right)\frac{\kappa'_{1b}}{e_{k,\mathrm{fd}}} + \left(S_3\zeta_k + S'_3\zeta'_k\right)\frac{\kappa'_{2b}}{e_{k,\mathrm{fd}}} \\
&\quad + \left(S_9 + S''_9\right)\left[\kappa'_{1b} + (\partial_{r_k} C_{\mathrm{fd}})\kappa'_{2b}\right]\frac{1}{e_{k,\mathrm{fd}}} \\
&\quad + \left(S_8\alpha_k + S'_8\alpha'_k\right)\frac{\kappa'_{1b}}{e_{k,\mathrm{fd}}} + \left(S_7\alpha_k + S'_7\alpha'_k\right)\frac{\kappa'_{2b}}{e_{k,\mathrm{fd}}} \\
&\quad + \frac{\kappa'_{2c}}{e_{k,\mathrm{fd}}}\left(\alpha_k\zeta_k + \alpha'_k\zeta'_k\right)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
J_{1,c,\mathrm{fd},4} &= -\frac{gz(\rho)}{2}\sum_{i,j,l}(\partial_{r_i}\partial_{r_j}\partial_{r_l} D_\rho)\frac{1}{\Delta^2}\left(\alpha_i\zeta_j\alpha_l\zeta_k + \alpha_i\zeta_j\alpha'_l\zeta'_k + \alpha'_i\zeta'_j\alpha_l\zeta_k + \alpha'_i\zeta'_j\alpha'_l\zeta'_k\right) \\
&= -\frac{gz(\rho)}{2\Delta^2}\Bigg\{\left(S_3S_7S_3\zeta_k + S_3S_7S'_3\zeta'_k + S'_3S'_7S_3\zeta_k + S'_3S'_7S'_3\zeta'_k\right)\kappa'_3
\end{aligned}
$$

$$
\begin{aligned}
&+\Big[(S_4S_7S_3\zeta_k + S_4S_7S_3'\zeta_k' + S_4'S_7'S_3\zeta_k + S_4'S_7'S_3'\zeta_k') \\
&+(S_3S_8S_3\zeta_k + S_3S_8S_3'\zeta_k' + S_3'S_8'S_3\zeta_k + S_3'S_8'S_3'\zeta_k') \\
&+(S_3S_7S_4\zeta_k + S_3S_7S_4'\zeta_k' + S_3'S_7'S_4\zeta_k + S_3'S_7'S_4'\zeta_k')\Big]\kappa_{2a}' \\
&+\Big[(S_4S_8S_3\zeta_k + S_4S_8S_3'\zeta_k' + S_4'S_8'S_3\zeta_k + S_4'S_8'S_3'\zeta_k') \\
&+(S_4S_7S_4\zeta_k + S_4S_7S_4'\zeta_k' + S_4'S_7'S_4\zeta_k + S_4'S_7'S_4'\zeta_k') \\
&+(S_3S_8S_4\zeta_k + S_3S_8S_4'\zeta_k' + S_3'S_8'S_4\zeta_k + S_3'S_8'S_4'\zeta_k')\Big]\kappa_{1a}' \\
&+(S_4S_8S_4\zeta_k + S_4S_8S_4'\zeta_k' + S_4'S_8'S_4\zeta_k + S_4'S_8'S_4'\zeta_k')\,\kappa_0' \\
&+(S_9S_4\zeta_k + S_9S_4'\zeta_k' + S_9''S_4\zeta_k + S_9''S_4'\zeta_k')\,\kappa_{1b}' \\
&+(S_9S_3\zeta_k + S_9S_3'\zeta_k' + S_9''S_3\zeta_k + S_9''S_3'\zeta_k')\,\kappa_{2b}' \\
&+(S_{10}S_8\zeta_k + S_{10}'S_8\zeta_k' + S_{10}'S_8'\zeta_k + S_{10}''S_8'\zeta_k')\,\kappa_{1b}' \\
&+(S_{10}S_7\zeta_k + S_{10}'S_7\zeta_k' + S_{10}'S_7'\zeta_k + S_{10}''S_7'\zeta_k')\,\kappa_{2b}' \\
&+\left(S_4S_9\zeta_k + S_4S_9'^a\zeta_k' + S_4'S_9'^b\zeta_k + S_4'S_9''\zeta_k'\right)\kappa_{1b}' \\
&+\left(S_3S_9\zeta_k + S_3S_9'^a\zeta_k' + S_3'S_9'^b\zeta_k + S_3'S_9''\zeta_k'\right)\kappa_{2b}' \\
&+(S_{11}\zeta_k + S_{11}'\zeta_k' + S_{11}''\zeta_k + S_{11}'''\zeta_k')\,\kappa_{2c}'\Big\}.
\end{aligned}
$$

Note the k dependence of the $J_{1,c,m}$ quantities through terms such as $e_{k,fd}$. Finally,

$$J_{1,c,fd} = J_{1,c,fd,1} + J_{1,c,fd,2} + J_{1,c,fd,3} + J_{1,c,fd,4}. \tag{75}$$

This completes the calculation for Approach 3.

25 Appendix–$J_{1,d}$ Formulae for Approaches 2 and 3

In this appendix expressions are computed for $J_{1,d}$ for Approaches 2 and 3. Similar to $J_{1,c}$, because $J_{1,d}$ is a contribution to $E(\Phi_k|y)$, there is a k dependence for $J_{1,d}$ that is suppressed in the notation. These expressions are required for the second order terms in the asymptotic expansions. Because second order terms are only required when $\rho_k = 0$, only that case is

considered..

The quantity $J_{1,d}$ is defined in Eq. 83 which, in terms of the standard notation, is

$$J_{1,d} = \frac{1}{2}\mathrm{tr}\left(L_\rho^{-1}\,\partial_{r_i}\partial_{r_j}\left[g_k(F_\rho(\bar{r}))\right]\Big|_{\bar{\rho}}\right).$$

Taking advantage of $\rho_k = 0$, this simplifies to

$$J_{1,d} = \frac{1}{2}\mathrm{tr}\left(L_\rho^{-1}Q_{\rho,k}\right). \tag{76}$$

First, consider C_{uc}. Define

$$\eta = \frac{1}{\sum_{k=1}^{\frac{L-1}{2}}\frac{\rho_k^2}{e_{k,\mathrm{uc}}} + R}$$

and use Eqs. 45 and 70 in Eq. 76 to compute that

$$\begin{aligned} J_{1,d,\mathrm{uc}} &= \frac{1}{2}\mathrm{tr}\left(Q_{\rho,k,\mathrm{uc}}L_{\rho,\mathrm{uc}}^{-1}\right) \\ &= g_Z(\rho)\left[\frac{\omega_{k,\mathrm{uc}}(\rho_0,\rho_k)}{e_{k,\mathrm{uc}}} - \eta\gamma_{k,\mathrm{uc}}\sum_{l=1}^{\frac{L-1}{2}}\omega_{l,\mathrm{uc}}(\rho_0,\rho_l)\gamma_{l,\mathrm{uc}}\right]. \end{aligned} \tag{77}$$

Since $\rho_k = 0$ by assumption, Eq. 46 simplifies to $\gamma_{k,\mathrm{uc}} = 0$ and Eq. 66 simplifies to

$$\omega_{k,\mathrm{uc}}(\rho_0,\rho_k) = b_{k,1c}.$$

Therefore, using Eq. 44, Eq. 77 simplifies to

$$\begin{aligned} J_{1,d,\mathrm{uc}} &= g_Z(\rho)\frac{b_{k,1c}}{e_{k,\mathrm{uc}}} \\ &= g_Z(\rho)\frac{b_{k,1c}}{f_k(0)+\tau\frac{4}{L}}. \end{aligned} \tag{78}$$

In this equation, $g_Z(\rho)$ depends on Ψ_l for all l, $b_{k,1c}$ depends on Ψ_k, and $f_k(0)$ and τ are independent of Ψ.

Second, consider C_{fd}. Use Eqs. 48 and 70 to compute that

$$\begin{aligned} J_{1,d,\mathrm{fd}} &= \frac{1}{2}\mathrm{tr}\left(Q_{\rho,k,\mathrm{fd}}L_{\rho,\mathrm{fd}}^{-1}\right) \\ &= \frac{1}{2}g_Z(\rho)\left[2\frac{\omega_{k,\mathrm{fd}}(\bar{\rho})}{e_{k,\mathrm{fd}}} - \frac{\zeta_k}{\Delta}\sum_{l=1}^{\frac{L-1}{2}}\omega_{l,\mathrm{fd}}(\bar{\rho})\alpha_l - \frac{\zeta_k'}{\Delta}\sum_{l=1}^{\frac{L-1}{2}}\omega_{l,\mathrm{fd}}(\bar{\rho})\alpha_l'\right. \\ &\quad \left. - \frac{\alpha_k}{\Delta}\sum_{l=1}^{\frac{L-1}{2}}\omega_{l,\mathrm{fd}}(\bar{\rho})\zeta_l - \frac{\alpha_k'}{\Delta}\sum_{l=1}^{\frac{L-1}{2}}\omega_{l,\mathrm{fd}}(\bar{\rho})\zeta_l'\right]. \end{aligned} \tag{79}$$

In this equation, $g_Z(\rho)$ depends on Ψ_l for all l, $\omega_{k,fd}$ depends on Ψ_0 and Ψ_k, and α_l, α'_l, ζ_l, ζ'_l, $e_{k,fd}$, and A are independent of Ψ. Unlike the C_{uc} case, the assumption that $\rho_k = 0$ does not dramatically simplify the equation. Note, however, that the l summations are linear in the size of the lattice and independent of k (so they need only be computed one per reconstruction). Therefore this is a practical computation.

26 Appendix–Multivariable second order asymptotic expansion formulae

In this appendix the second order asymptotic expansion is computed by Laplace's method of the n-dimensional integral

$$I(\lambda) = \int_D q(x) \exp(\lambda r(x)) \mathrm{d}x^n$$

when $r(x)$ has a single global maximum which is located in the interior of the region D C El". The location, denoted p, of the global maximum is the critical point.

The plan, as described for the scalar case in Ref. [14, pp. 272-2741, has four steps:

1. Express q and r by Taylor series expansions around the critical point p.
2. Express the exponential of the Taylor series of r by the product of two terms: The first term is the exponential of the first three terms in the Taylor series expansion of r. The second term is the Taylor series expansion around 0 of the exponential function, evaluated at the sum of the fourth and higher order terms in the Taylor series expansion of r.
3. Collect terms of the same order in the asymptotic parameter A.
4. Finally, approximate the region D by El" and exactly evaluate the resulting Gaussian integrals.

In order to compute a second order asymptotic expansion of the integral it turns out that Taylor series terms up to order 4 in r and order 2 in q must be accountecl for.

Let

$$q_{i,j,\ldots,m}(x) = \frac{\partial^M q}{\partial x_i \partial x_j \cdots \partial x_m}(x)$$
$$q_{i,j,\ldots,m} = q_{i,j,\ldots,m}(\rho)$$

where M is the total number of indices $i, j, \ldots, m$. Use corresponding notation for $r(x)$. Note the important fact that $r_i = 0$ for all i because the critical point is in the interior of D. Let $\bar{\bar{q}}$ (F) be the Hessian matrix for q (r) with entries $q_{i,j}$ $(r_{i,j})$ and define $R = -\bar{\bar{r}}^{-1}$ with entries $R_{i,j}$. Sums over indices always range from 0 to $n-1$ for each index.

Define $s = \sqrt{\lambda}(x - p)$ with components s_i, $i \in \{0, \ldots, n-1\}$. Then the Taylor series around p of q and r are

$$q(x) = q + \frac{1}{\sqrt{\lambda}} \sum_i q_i s_i + \frac{1}{2\lambda} \sum_{i,j} q_{i,j} s_i s_j + \cdots$$

and

$$r(x) = r + \frac{1}{\sqrt{\lambda}} \sum_i r_i s_i + \frac{1}{2\lambda} \sum_{i,j} r_{i,j} s_i s_j + \frac{1}{3!\lambda^{3/2}} \sum_{i,j,k} r_{i,j,k} s_i s_j s_k + \frac{1}{4!\lambda^2} \sum_{i,j,k,l} r_{i,j,k,l} s_i s_j s_k s_l + \ldots.$$

The previously described plan leads to the following series of equations:

$$\begin{aligned}
I(\lambda) &= \int_D q(x) \exp(\lambda r(x)) dx^n \\
&= \int \left(q + \frac{1}{\sqrt{\lambda}} \sum_i q_i s_i + \frac{1}{2\lambda} \sum_{i,j} q_{i,j} s_i s_j + \cdots \right) \exp\left(\lambda \left(r + \frac{1}{\sqrt{\lambda}} \sum_i r_i s_i + \frac{1}{2\lambda} \sum_{i,j} r_{i,j} s_i s_j \right.\right. \\
&\quad \left.\left. + \frac{1}{3!\lambda^{3/2}} \sum_{i,j,k} r_{i,j,k} s_i s_j s_k + \frac{1}{4!\lambda^2} \sum_{i,j,k,l} r_{i,j,k,l} s_i s_j s_k s_l + \ldots \right) \right) ds^n / \lambda^{n/2} \\
&= \int \left(q + \frac{1}{\sqrt{\lambda}} \sum_i q_i s_i + \frac{1}{2\lambda} \sum_{i,j} q_{i,j} s_i s_j + \cdots \right) \exp\left(\lambda \left(r + \frac{1}{\sqrt{\lambda}} \sum_i r_i s_i + \frac{1}{2\lambda} \sum_{i,j} r_{i,j} s_i s_j \right) \right) \times \\
&\quad \times \left[1 + \lambda \left(\frac{1}{3!\lambda^{3/2}} \sum_{i,j,k} r_{i,j,k} s_i s_j s_k + \frac{1}{4!\lambda^2} \sum_{i,j,k,l} r_{i,j,k,l} s_i s_j s_k s_l + \cdots \right) \right.
\end{aligned}$$

$$+\frac{\lambda^2}{2}\left(\frac{1}{3!\lambda^{3/2}}\sum_{i,j,k} r_{i,j,k}s_is_js_k+\frac{1}{4!\lambda^2}\sum_{i,j,k,l} r_{i,j,k,l}s_is_js_ks_l+\cdots\right)^2+\cdots\Big]ds^n/\lambda^{n/2}$$

$$= \int\left(q+\frac{1}{\sqrt{\lambda}}\sum_i q_is_i+\frac{1}{2\lambda}\sum_{i,j} q_{i,j}s_is_j+\cdots\right)\exp\left(\lambda r+\sqrt{\lambda}\sum_i r_is_i+\frac{1}{2}\sum_{i,j} r_{i,j}s_is_j\right)\times$$

$$\times\left\{1+\left(\frac{1}{3!\sqrt{\lambda}}\sum_{i,j,k} r_{i,j,k}s_is_js_k+\frac{1}{4!\lambda}\sum_{i,j,k,l} r_{i,j,k,l}s_is_js_ks_l+\cdots\right)\right.$$

$$\left.+\left[\frac{1}{2(3!)^2\lambda}\left(\sum_{i,j,k} r_{i,j,k}s_is_js_k\right)^2+\cdots\right]+\cdots\right\}ds^n/\lambda^{n/2}$$

$$= \int\frac{1}{\lambda^{n/2}}\exp\left(\lambda r+\sqrt{\lambda}\sum_i r_is_i+\frac{1}{2}\sum_{i,j} r_{i,j}s_is_j\right)\Big)\Big\{q+\frac{1}{\sqrt{\lambda}}\left[\frac{q}{3!}\sum_{i,j,k} r_{i,j,k}s_is_js_k+\sum_i q_is_i\right]$$

$$+\frac{1}{\lambda}\left[\frac{q}{4!}\sum_{i,j,k,l} r_{i,j,k,l}s_is_js_ks_l+\frac{q}{2(3!)^2}\left(\sum_{i,j,k} r_{i,j,k}s_is_js_k\right)^2+\left(\sum_i q_is_i\right)\left(\frac{1}{3!}\sum_{i,j,k} r_{i,j,k}s_i\right)\right.$$

$$\left.+\frac{1}{2}\sum_{i,j} q_{i,j}s_is_j\right]+\cdots\Big\}ds^n$$

$$= N^{-1}(-\bar{\bar{r}})\frac{1}{\lambda^{n/2}}\exp(\lambda r)\int N(-\bar{\bar{r}})\exp[-s^T(-\bar{\bar{r}})s]\Big\{q+\frac{1}{\sqrt{\lambda}}\left[\frac{q}{3!}\sum_{i,j,k} r_{i,j,k}s_is_js_k+\sum_i q_is_i\right]$$

$$+\frac{1}{\lambda}\left[\frac{q}{4!}\sum_{i,j,k,l} r_{i,j,k,l}s_is_js_ks_l+\frac{q}{2(3!)^2}\left(\sum_{i,j,k} r_{i,j,k}s_is_js_k\right)^2+\left(\sum_i q_is_i\right)\left(\frac{1}{3!}\sum_{i,j,k} r_{i,j,k}s_i\right)\right.$$

$$\left.+\frac{1}{2}\sum_{i,j} q_{i,j}s_is_j\right]+\cdots\Big\}ds^n$$

$$= N^{-1}(-\bar{\bar{r}})\frac{1}{\lambda^{n/2}}\exp(\lambda r)\left[J_{0,a}+\frac{1}{\sqrt{\lambda}}\left(J_{1/2,a}+J_{1/2,b}\right)+\frac{1}{\lambda}\left(J_{1,a}+J_{1,b}+J_{1,c}+J_{1,d}\right)+\cdots\right]$$

$$= N^{-1}(-\lambda\bar{\bar{r}})\exp(\lambda r)\left[J_{0,a}+\frac{1}{\sqrt{\lambda}}\left(J_{1/2,a}+J_{1/2,b}\right)+\frac{1}{\lambda}\left(J_{1,a}+J_{1,b}+J_{1,c}+J_{1,d}\right)+\cdots\right].$$

The zero order term is $J_{0,a}$, the first order terms are $\frac{1}{\sqrt{\lambda}}\left(J_{1/2,a}+J_{1/2,b}\right)$, and the second order terms are $\frac{1}{\lambda}\left(J_{1,a}+J_{1,b}+J_{1,c}+J_{1,d}\right)$.

Each of the $J_{\nu,s}$ is related to moments of a zero mean Gaussian random vector with covariance matrix $-\bar{\bar{r}}$. These integrals can be evaluated in two steps:

1. Apply Gaussian moment factoring [17, p. 229].

2. Simplify the results by symmetry arguments based on the equality of mixed partial derivatives that differ only in the order of the derivatives. For instance, $r_{i,j,k} = r_{i,k,j} =$

$r_{j,i,k} = r_{k,i,j} = r_{j,k,i} = r_{k,j,i}$ for any choice of i, j, and k. In particular, $r_{i,j} = r_{j,i}$ and therefore $R_{i,j} = R_{j,i}$.

Therefore,

$$
\begin{aligned}
J_{0,a} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] q \mathrm{d}s^n \\
&= q, \\
J_{1/2,a} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] \frac{q}{3!} \sum_{i,j,k} r_{i,j,k} s_i s_j s_k \mathrm{d}s^n \\
&= 0, \\
J_{1/2,b} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] \sum_i q_i s_i \mathrm{d}s^n \\
&= 0, \\
J_{1,a} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] \frac{q}{4!} \sum_{i,j,k,l} r_{i,j,k,l} s_i s_j s_k s_l \mathrm{d}s^n \\
&= \frac{q}{4!} \sum_{i,j,k,l} r_{i,j,k,l} (R_{i,j}R_{k,l} + R_{i,k}R_{j,l} + R_{i,l}R_{j,k}) \\
&= \frac{q}{8} \sum_{i,j,k,l} r_{i,j,k,l} R_{i,j} R_{k,l}, \qquad (80) \\
J_{1,b} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] \frac{q}{2(3!)^2} \left(\sum_{i,j,k} r_{i,j,k} s_i s_j s_k \right)^2 \mathrm{d}s^n \\
&= \frac{q}{2(3!)^2} \sum_{i,j,k,i',j',k'} r_{i,j,k} r_{i',j',k'} (R_{i,j}R_{k,i'}R_{j',k'} + R_{i,j}R_{k,j'}R_{i',k'} + R_{i,j}R_{k,k'}R_{i',j'} \\
&\quad + R_{i,k}R_{j,i'}R_{j',k'} + R_{i,k}R_{j,j'}R_{i',k'} + R_{i,k}R_{j,k'}R_{i',j'} + R_{i,i'}R_{j,k}R_{j',k'} \\
&\quad + R_{i,i'}R_{j,j'}R_{k,k'} + R_{i,i'}R_{j,k'}R_{k,j'} + R_{i,j'}R_{j,k}R_{i',k'} + R_{i,j'}R_{j,i'}R_{k,k'} \\
&\quad + R_{i,j'}R_{j,k'}R_{k,i'} + R_{i,k'}R_{j,k}R_{i',j'} + R_{i,k'}R_{j,i'}R_{k,j'} + R_{i,k'}R_{j,j'}R_{k,i'}) \\
&= \frac{q}{2(3!)^2} \sum_{i,j,k,i',j',k'} r_{i,j,k} r_{i',j',k'} (9R_{i,j}R_{k,k'}R_{i',j'} + 6R_{i,i'}R_{j,j'}R_{k,k'}) \\
&= q \left[\frac{1}{8} \sum_{k,k'} \left(\sum_{i,j} r_{i,j,k} R_{i,j} \right) \left(\sum_{i',j'} r_{i',j',k'} R_{i',j'} \right) R_{k,k'} \right. \\
&\quad \left. + \frac{1}{12} \sum_{i,j,k,i',j',k'} r_{i,j,k} r_{i',j',k'} R_{i,i'} R_{j,j'} R_{k,k'} \right], \qquad (81) \\
J_{1,c} &= \int N(-\bar{\bar{r}}) \exp[-s^T(-\bar{\bar{r}})s] \left(\sum_i q_i s_i \right) \left(\frac{1}{3!} \sum_{i,j,k} r_{i,j,k} s_i \right) \mathrm{d}s^n
\end{aligned}
$$

$$= \frac{1}{3!}\sum_{i,j,k,l} r_{i,j,k}q_l(R_{i,j}R_{k,l} + R_{i,k}R_{j,l} + R_{i,l}R_{j,k})$$

$$= \frac{1}{2}\sum_{i,j,k,l} r_{i,j,k}q_l R_{i,j}R_{k,l}, \tag{82}$$

and $$J_{1,d} = \int N(-\bar{\bar{r}})\exp[-s^T(-\bar{\bar{r}})s]\frac{1}{2}\sum_{i,j} q_{i,j}s_i s_j ds^n$$

$$= \frac{1}{2}\sum_{i,j} q_{i,j}R_{i,j}$$

$$= \frac{1}{2}\mathrm{tr}(\bar{\bar{q}}R). \tag{83}$$

Therefore the final result is

$$I(\lambda) = N^{-1}(-\lambda\bar{\bar{r}})\exp(\lambda r)\left[q + \frac{1}{\lambda}(J_{1,a} + J_{1,b} + J_{1,c} + J_{1,d}) + \cdots\right]. \tag{84}$$

Note that a particular coefficient from the Taylor series, say the n-th order coefficient, appears in multiple terms of the asymptotic expansion (essentially the n-th and higher order terms). For example, q_i, the first order coefficient in the q Taylor series, would appear in the first order term of the asymptotic expansion (i.e., the $\frac{1}{\sqrt{\lambda}}$ term $J_{1/2,b}$), except that both the first order terms are zero by symmetry, and does appear in the second order term of the asymptotic expansion (i.e., the $\frac{1}{\lambda}$ term $J_{1,c}$).

References

[1] Peter C. Doerschuk. Bayesian signal reconstruction, Markov random fields, and x-ray crystallography. Journal of the Optical Society of America A, 8(8):1207–1221, 1991.

[2] 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, 8(8):1222–1232, 1991.

[3] Paul B. Yale. Geometry and Symmetry. Dover Publications, Inc., New York, 1988. originally published by Holden-Day, San Francisco, 1968.

[4] Theo Hahn, editor. *International Tables for X-Ray Crystallography, Volume A: Space Group Symmetry*. D. Reidel Publishing Company, Dordrecht-Boston-Lancaster-Tokyo, second revised edition', 1987.

[5] J. J. Burckhardt. *Die Bewegungsgruppe der Kristallographie.* Birkhauser, Basel, 1957.

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

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

[8] 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.

[9] Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6(6):721–741, November 1984.

[10] John K. Goutsias. Mutually compatible Gibbs random fields. *IEEE Transactions on Information Theory,* 35(6), November 1989.

[11] George H. Stout and Lyle H. Jensen. *X-ray Structure Determination: A Practical Guide.* Macmillan Publishing Co., Inc., New York, 1968.

[12] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1988.

[13] The MathWorks, Inc. *PRO-MATLAB User's Guide.* The MathWorks, Inc., 21 Eliot Street, South Natick, MA 01760, 1990.

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

[15] Alexander Morgan. ***Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems.*** Prentice-Hall, Englewood Cliffs, N.J., 1987.

[16] IMSL, Inc. ***Users Manual, IMSL MATH/LIBRARY.*** IMSL, Inc., 2500 ParkWest Tower One!,2500 City West Boulevard, Houston, TX 77042, December 1989. Version 1.1.

[17] Harry L. van Trees. ***Detection, Estimation, and Modulation Theory: Part I.*** John Wiley and Sons, Inc., New York, 1968.