Adaptive Bayesian Signal Reconstruction with A Priori Model Implementation and Synthetic Examples for X-ray Crystallography

Peter C. Doerschuk

School of Electrical Engineering

Purdue University

West Lafayette, Indiana 47907–1285

(317) 494–1742

February 26, 1991

Abstract

In [1] a signal reconstruction problem motivated by X-ray crystallography is (approximately) solved in a Bayesian statistical approach. The signal is zero-one, periodic, and substantial statistical a priori information is known, which is modeled with a Markov random field. The data are inaccurate magnitudes of the Fourier coefficients of the signal. The solution is explicit and the computational burden is independent of the signal dimension. In this paper a detailed parameterization of the a priori model appropriate for crystallography is proposed and symmetry-breaking parameters in the solution are used to perform data-dependent adaptation of the estimator. The adap-

tation attempts to minimize the effects of the spherical model approximation used in the solution. Several examples in one and two dimensions 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*, 1991. To appear.

1 Introduction

In [1] a novel Bayesian statistical approach is presented to a class of phase-retrieval problems exemplified by the inverse problem of single-crystal X-ray crystallography. In this paper parameters are proposed in the a priori model that are suitable for modeling bond-length limitations in covalently-bound molecules, free parameters in the estimator are employed in order to design a data-dependent adaptive estimator, and several numerical examples in one and two dimensions based on simulated data are provided.

The situation from [1] is recalled in this and the following paragraph. There is a periodic object which takes only values zero and one, and the observer makes corrupted measurements, denoted y_k, of the magnitude of its 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 independent Gaussian random variables with k-dependent variances. Both the MRF and the Gaussian observation errors can be put in the form of energy functions with the corresponding probabilities in the form of the Gibbs distribution. A Bayesian conditional mean estimation problem is approximately solved. 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. If one blindly averages over all configurations of $\{\phi\}$, the result is a DC field. Therefore, an additional term in the energy function 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. The ψ_n are the free parameters alluded to above, and the subject of this paper is a method

for choosing them in a data-dependent fashion.

In solving the estimation problem in [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. (Actually two different asymptotic problems were proposed and solved depending on whether the a priori energy function was scaled along with the observational energy function). 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.

For any choice of ψ this is a valid Bayesian estimation problem. Therefore, the goal is to choose ψ in an intelligent fashion. One natural approach, which is the approach taken here, is to try to choose ψ, or equivalently, the estimation problem, to minimize the effects of the two approximations made in the solution of the estimation problem.

In this paragraph the Hamiltonian is recalled from [1]. The Hamiltonian is in three parts—the a priori probability part H^{apriori}, the conditional observational probability part H^{obs}, and the symmetry breaking part $H^{\text{s.b.}}$. That is,

$$H = H^{\text{apriori}} + H^{\text{obs}} + H^{\text{s.b.}}.$$

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,

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

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

where L is the size of the lattice which is also the period of the crystal when measured in

4

lattice spacings. The conditional observational part is Gaussian, specifically,

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

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

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

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 with the kernel function ψ, specifically,

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

where ψ_n is real and periodic with period L.

In this paragraph the asymptotics are recalled from [1]. Two different asymptotic limits are considered. The first limit, denoted Problem 1, is purely a small observation noise limit. That is, the observation noise variance σ_k^2 satisfies $\sigma_k^2 = \frac{1}{\lambda} \bar{\sigma}_k^2$ and $\lambda \uparrow \infty$. The second limit, denoted Problem 2, combines the small observation noise limit with a proportional scaling of the a priori Hamiltonian. That is, the observation noise variance σ_k^2 and the parameters $W_2(k_1, k_2)$ and w_1 from the a priori portion of the Hamiltonian satisfy $\sigma_k^2 = \frac{1}{\lambda} \bar{\sigma}_k^2$, $W_2(k_1, k_2) = \lambda \chi \bar{W}_2(k_1, k_2)$, $w_1 = \lambda \chi \bar{w}_1$, $\lambda \uparrow \infty$, and χ is a fixed real number.

The remainder of this paper is organized in the following fashion. In Section 2 a parameterization of the a priori Hamiltonian is described that is appropriate for X-ray crystallography. In Section 3 an optimization problem is described for the choice of ψ. Implementation issues are briefly discussed in Section 4. In Section 5 one-dimensional examples are presented which emphasize the need to break the symmetries discussed above. In Section 6 a one-dimensional example is presented which is sufficiently small such that it is possible to compute estimator performance statistics versus observation noise intensity for the exact conditional mean estimator, the approximations to the conditional mean estimator (with either definition of the asymptotics), and the trivial estimator that simply takes a random

sample from the a priori distribution for its estimate. A two-dimensional example is presented in Section 7. Finally, in Section 8 the results to date are discussed and directions for future research are described.

2 Use of H^{apriori} in X-Ray Crystallography

In this section one method for using the quadratic a priori Hamiltonian H^{apriori} to model bond-length limitations in covalently bound molecules is described. By interpreting $|\cdot|$ as absolute value in one dimension or as Euclidean norm in higher dimensions, the following applies for arbitrary dimension of the signal.

Define a minimum (maximum) bond length l_1 (l_2) measured in lattice spacings. Assume that there is an atom at site n_0. Therefore, $\phi_{n_0} = 1$. In order to avoid the presence of a neighboring atom at less than the minimum bond length, the a priori probability function is low if there is an atom in the region $\{n||n - n_0| < l_1\}$. In addition, in order to avoid unbound atoms, the a priori probability function is low if there is no neighboring atom in the region $\{n|l_1 \leq |n - n_0| < l_2\}$. In order to use these ideas with the quadratic H^{apriori} defined previously, the penalty on crowding (i.e., an atom in the region $\{n||n - n_0| < l_1\}$) must become progressively greater as more atoms fall in the region and the penalty on being unbound (i.e., $\{n|l_1 \leq |n - n_0| < l_2\}$) must become progressively smaller as more atoms fall in the region. (Note, however, that as more atoms are present in this region they in turn will become crowded with respect to each other, thereby lowering the a priori probability function again). The penalties described here are associated with site n_0 and therefore w_1, w_2 should be chosen so that the penalties appear in u_{n_0} (eqn. 1). Since there are no penalties that depend only on the value of a single site, $w_1 = 0$. Since the penalties only occur when $\phi_{n_0} = 1$, it would be convenient to choose $w_2(n_1, n_2)$ so that it is zero unless $n_1 = 0$. Since w_2 is symmetric, define an asymmetric $\tilde{w}_2$ whose symmetric part is w_2. The definitions are

$w_2(n_1, n_2) = \frac{1}{2}(\tilde{w}_2(n_1, n_2) + \tilde{w}_2(n_2, n_1))$ where $\tilde{w}_2(n_1, n_2)$ is defined by

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

$$\tilde{w}_2(0, n_2) = \begin{cases} p_1 & \text{if } 1 \leq |n_2| < l_1 \\ p_2 & \text{if } l_1 \leq |n_2| < l_2 \\ 0 & \text{otherwise} \end{cases}$$

and where $p_1 > 0$, $p_2 < 0$, and $1 \leq l_1 \leq l_2$. Generalizations to Hamiltonians that include a quadratic penalty on deviations from a desired valence number are also straightforward.

To this point in the present paper and throughout [1], this model and estimator have been motivated from the perspective of atomic resolution X-ray crystallography where the lattice variables are the presence or absence of an atom at that location. However, there are low resolution structures of large molecules (e.g., proteins) that might also be addressed. In the low resolution protein setting, one is interested in locating the path in three dimensions of a solid "tube" of high electron density that follows the polypeptide backbone in a background of relatively low electron density. The new interpretation of the lattice variables would then be that 0 corresponds to background and 1 corresponds to polypeptide backbone, which is approximated to have only one value of electron density. The purpose of the a priori Hamiltonian H^{apriori} would then be quite different. Specifically, it would seek to make the occupied sites (the 1's) form a connected solid tube rather than scattering through space. Therefore, it would more resemble a ferromagnetic Ising model Hamiltonian than the bond-length Hamiltonian described in the previous paragraphs. After this change in interpretation of the lattice variables and change in the constants in the a priori Hamiltonian, the remainder of the calculation in [1] would be unaltered.

3 The Choice of ψ

The symmetry breaking portion of the Hamiltonian, $H^{\text{s.b.}}$, contains the parameters ψ_n and for any choice of ψ_n an approximation to the Bayesian conditional mean estimator was computed in [1]. These degrees of freedom are used to design an adaptive estimator.

As described in the Introduction, the calculations of [1] contain two approximations, namely the spherical and asymptotic approximations. The asymptotic approximation is of less concern because it can be improved on by taking additional terms in the asymptotic series and because the crystallography problem is in fact a small-noise problem. However, the spherical approximation, by relaxing the atomicity of the electron density which has been central to successful direct methods in crystallography, is of more concern. In addition, natural improvements on the spherical approximation are difficult to compute because they couple the Fourier coefficients. Therefore, the adaptive estimation scheme focuses on decreasing the effects of the spherical approximation.

The idea is to choose a new ψ for each data set such that the choice minimizes a cost function C. C is a function both of ψ_n and of the unthresholded approximations m_n to the conditional expectations $E(\phi_n|\{y\})$ that result from using ψ_n in the estimator. The cost function is a linear combination $C = \gamma_1 C_1 + \gamma_2 C_2 + \gamma_3 C_3$ of three individual costs. C_1 attempts to improve on the spherical model by penalizing a manifestation of the spherical model's relaxation of the 0-1 constraints on the lattice variables. C_2 attempts to control the energy in ψ_n so that it neither vanishes and thereby fails to break the symmetries nor dominates the Hamiltonian. C_3 attempts to select an estimator that is confident of its answer, that is, an estimator where m_n are near 0 or 1.

For binary lattice variables the conditional expectations are constrained to $[0, 1]$. The estimates computed using the spherical approximation routinely violate this condition. Therefore, if m_n is the computed approximation to $E(\phi_n|\{y\})$, ψ is chosen to minimize a cost function that penalizes $m_n \notin [0, 1]$. The work described here uses the simple cost function

$$
\begin{aligned}
C_1 &= \sum \sigma_1(m_n) \\
\sigma_1(x) &= \begin{cases} (x-1)^2 & \text{if } x > 1 \\ 0 & \text{if } 0 \le x \le 1 \\ x^2 & \text{if } x < 0 \end{cases} \cdot
\end{aligned}
$$

At the same time, it is undesirable for ψ to vanish because of its role in symmetry breaking,

or to grow too large and dominate the estimator. Therefore, deviations in energy from a desired energy E_0 are penalized with the cost function

$$C_2 = (||\psi||_2^2 - E_0)^2$$

where $||x||_p = (\sum_n x_n^p)^{1/p}$. Choice of E_0 is discussed in the following section.

Finally, among all the estimators parameterized by ψ, it is desirable to choose a confident estimator, that is, an estimator for which the resulting m_n are near 0 and 1. Therefore deviations in m_n from 0 and 1 are penalized by using the cost function

$$
\begin{aligned}
C_3 &= \sum \sigma_3(m_n) \\
\sigma_3(x) &= \begin{cases} 0 & \text{if } x > 1 \\ (x-1)^2 x^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{if } x < 0 \end{cases} .
\end{aligned}
$$

(I prefer to penalize $m_n > 1$ and $m_n < 0$ through C_1 only).

The costs C_1 and C_3 depend on m_n, the approximation to the expectation $E(\phi_n|\{y\})$, which in turn is a function of ψ. When optimizing ψ the m_n are recomputed for each new value of ψ.

There are obviously many other ways in which one might set ψ. One natural entirely different category of methods would be to set ψ so that the resulting Bayesian estimator has some desirable statistical property. Here the choice would be made once forever according to some ensemble criteria while in the method described above the choice is made for each data set according to a pathwise criteria. For any ψ, the Bayesian conditional mean estimator minimizes the conditional expectation of the squared difference between the actual and estimated fields. One might then choose ψ to minimize higher order statistics or a special discrete class of errors that are particularly important from the application perspective.

4 Implementation of the ψ Optimization

When computing ψ, globally optimal solutions were not sought. Rather I have thought in terms of downhill search starting from an initial condition $\psi^{i.c.}$ where the initial condition (1) has a clear symmetry-breaking interpretation and (2) is used to set E_0 in C_2 by $E_0 = ||\psi^{i.c.}||_2^2$. This choice seems preferable because it is computationally straightforward and because it attempts to keep some balance between the symmetry-breaking role of ψ, only implicitly included in the optimization criteria, and the optimization role of ψ (since a simple downhill optimization will presumably remain in some domain of attraction of the initial condition and the initial condition has a clear symmetry-breaking interpretation).

For one-dimensional problems the initial condition was $\psi_n^{i.c.} = \begin{cases} 0 & \text{if } n = 0 \\ (L-n)/L & \text{if } n \neq 0 \end{cases}$.
This choice sets $H^{s.b.}$ to be the first moment of the field ϕ_n and therefore clearly breaks translational symmetry. Two-dimensional problems have more types of symmetries than one-dimensional problems and an initial condition that clearly breaks more than translational symmetry is desirable. In the work described here the initial condition was $\psi_n^{i.c.} = \bar{\psi}(n_1; L_1, \alpha_1)\bar{\psi}(n_2; L_2, \alpha_2)$ where L_1 and L_2 are the dimensions of the lattice, α_1 and α_2 are constants, and

$$\bar{\psi}(n; L, \alpha) = \begin{cases} 0 & \text{if } n = 0 \\ (L-n)(1 + \alpha(L-n))/L & \text{if } n \neq 0 \end{cases}. \tag{2}$$

This choice breaks translational and inversion (equivalently, rotations by π) symmetries. If $L_1 = L_2$ then rotations by $n\frac{\pi}{2}$ are also symmetry operators and these symmetries are also broken by this choice of ψ.

Though the initial condition for ψ in two-dimensional problems is a product of functions of individual coordinates, the optimization considered any two-dimensional ψ not restricted to this product form. However, it may be sufficient and would certainly save computational expense to consider throughout the optimization only ψ with the product form.

A very inefficient scheme was used for optimizing ψ which, however, results in an unconstrained optimization problem. The inefficiency stems from the parameterization. Recall

from [1] that the estimator has the following structure. First, fix ψ. Second, compute approximations M_k to $E(\Phi_k|\{y\})$, the conditional expectations of the Fourier coefficients of the field ϕ_n. The formulae for this computation are in terms of Ψ_k, the Fourier coefficients of ψ_n and the data $\{y\}$. Third, compute approximations m_n to $E(\phi_n|\{y\})$, the conditional expectation of the field, by computing the inverse Fourier series of M_k. Fourth, threshold the m_n to compute the estimated field.

The L real variables ψ_n $n \in \{0, \ldots, L-1\}$ were used to parameterize ψ. Because each of these variables independently ranges over all of R, methods for unconstrained optimization problems can be applied. However, a computational penalty is paid because first ψ_n must be transformed into Ψ_k, at the cost of one Fourier series, and then M_k must be transformed into m_n, at the cost of a second Fourier series. If the parameterization was $\Psi_0 \in R$, $\angle \Psi_k$, and $|\Psi_k|$ $(k \in \{1, \ldots, \frac{L-1}{2}\})$, then the optimization is no longer unconstrained since $\angle \Psi_k \in [0, 2\pi)$ and $|\Psi_k| \in [0, \infty)$, but the cost of one of the Fourier series calculations is avoided. Note in this regard that C_2 can easily be computed in terms of ψ_n or in terms of $|\Psi_k|$. Also note that the above is purely a discussion of efficiency per iteration. The behavior of the optimization criteria as a function of the parameterization (e.g., straight versus highly curved valleys) will determine the number of iterations required and this aspect has not been considered at all.

Furthermore, an inefficient multidimensional downhill simplex method [2] was used to solve the unconstrained optimization problem. This method uses function evaluations but no derivatives and attempts to move a nondegenerate simplex in R^L defined originally by $L + 1$ initial condition vectors downhill by simple geometric moves (e.g., reflection of the vertex with the highest function value through the opposite face of the simplex to a position with a lower function value) until the simplex arrives at a minimum. This algorithm was chosen because the software is robust and was easily available. As mentioned in the previous section, at each step m_n (the approximation to $E(\phi_n|\{y\})$) is recomputed for each choice of ψ.

Finally, all the equations, both for the one- and two-dimensional cases, are for odd

dimensioned lattices. As mentioned in [1], this case was considered because the fact that ϕ_n is real constrains Φ_k, its Fourier coefficients. Specifically, in one dimension, it implies that $\Phi_k = \Phi^*_{L-k}$. For L even there are therefore two coefficients that are guaranteed real (more in higher dimensions) while for L odd there is one such coefficient in any dimension. Therefore the equations are much simpler for L odd. However, L odd is unfavorable from the computational point of view because it is incompatible with using power of two Fast Fourier Transforms. In one dimension the Fourier series was computed by explicit summation at a cost proportional to L^2 while in two dimensions the Fourier series was computed by a one-dimensional chirp z transform on rows and then on columns. In light·of the inefficient parameterization, software, and Fourier series situation, measures of algorithm speed are not reported.

In [1] a very brief sketch was given of an algorithm for locating the critical point in the asymptotics that is exact, noniterative, and requires only order L computation. This algorithm is described more fully in the Appendix. In the examples described in this paper, there is always a single critical point so the simple formulae of [1] apply. In fact, in the problems where I have investigated, there is only a single maximum for the exponent which determines the critical point.

5 One-Dimensional Examples

These examples are included to emphasize the importance of the symmetry breaking role of ψ.

The calculations of this section use the Problem 2 asymptotics where the observational Hamiltonian H^{obs} and a priori Hamiltonian H^{apriori} are both scaled by the asymptotic parameter and the ratio between their scalings is χ. In the Hamiltonian of Sections 1 and 2 take $\sigma_k^2 = \sigma^2 = 0.1 \ \forall k \in \{0, \ldots, L-1\}$, $\chi = 1$, $q = 1$, $p_1 = 1$, $p_2 = -2$, $l_1 = 4$, $l_2 = 6$, and $L = 27$. For the adaptive estimator design cost take either $C^0 = C_1 + C_2$ or $C^1 = C_1 + C_2 + C_3$. Two

12

simulations are described based on two different choices of the underlying field $\{\phi\}$. In both cases the simulated data are pseudo Gaussian random numbers with variance σ^2 added to the Fourier coefficients of the field. Since the fields have either two or three occupied sites, these noise levels correspond well to the 1 to 3 percent errors in real crystallographic data [3, p. 193].

For simulated data based on the field $\phi_n^2 = \begin{cases} 0 & \text{if } n \notin \{1,5\} \\ 1 & \text{if } n \in \{1,5\} \end{cases}$ using cost C^0 (respectively, C^1) the approximate conditional expectations are shown in Figure 1a (respectively, Figure 2a) and optimized ψ_n functions are shown in Figure 1b (respectively, Figure 2b). In both cases the thresholded estimate will be exactly correct though in the case of C^0 (i.e., not specifically selecting a confident estimator), the estimator almost fails to locate the first occupied site.

For simulated data based on the field $\phi_n^3 = \begin{cases} 0 & \text{if } n \notin \{1,5,9\} \\ 1 & \text{if } n \in \{1,5,9\} \end{cases}$ using cost C^0 (respectively, C^1) the approximate conditional expectations are shown in Figure 3a (respectively, Figure 4a) and optimized ψ_n functions are shown in Figure 3b (respectively, Figure 4b). In the case of C^0 (i.e., not specifically selecting a confident estimator), the estimator fails to locate the first and third occupied sites. Instead there are two broad low peaks. The reason for the broad low peaks is a failure to adequately break inversion symmetry. Specifically, the left (respectively, right) half of the left peak, the central peak, and the left (respectively, right) half of the right peak corresponds to intervals of 5 and 4 (respectively, 4 and 5) corresponding to the signal and the inverted signal. When C^1 is used, this inversion symmetry is broken and only one of the signal and its inversion (specifically its inversion) are substantially represented in the a posteriori expectation and the thresholded estimate is exactly correct.

6 Estimator Performance Statistics Versus Observation Noise Intensity

This section describes results concerning the estimator performance statistics versus observation noise intensity for the exact conditional mean estimator, the two approximations to the conditional mean estimator (that is, the two definitions of the asymptotics), and the trivial estimator that disregards the data and simply takes a random sample from the a priori distribution for its estimate. Comparison of the performance of the nontrivial and trivial estimators for different performance measures provides measures of the signal to noise ratio improvement due to the use of the nontrivial estimator. To perform these calculations the lattice must be very small–one-dimensional with $L = 17$ sites–because the computation is done by Monte Carlo methods and the calculation of an individual estimate for the exact estimator is done by exhaustive enumeration.

The data is generated by first computing configurations of the H^{apriori} described in Section 2 and then computing the observational transformation (Fourier transform followed by magnitude squared) and adding Gaussian pseudo-random variables. The combination of configuration generation, observational transformation, and additive noise is denoted the "forward" problem.

The three nontrivial inverse problems are matched to the forward problem, that is, they share the same values for the parameters, except that the configurations in the forward problem are chosen with $q = 0$ for the symmetry breaking parameter.

H^{apriori} has parameters $w_1 = 0$, $l_1 = 3$, $l_2 = 5$, $p_1 = 1.5$, and $p_2 = -0.5$. The observational transformation has no parameters. The additive white pseudo-random Gaussian noise is parameterized by its standard deviation, denoted σ_k, which is constant over k. A range of values for σ were considered. See the figures. In all calculations, there are no missing observations.

The exact conditional mean inverse problem is matched to the forward problem plus the

addition of $q = 1.0$.

The approximate conditional mean inverse problems use $\gamma_1 = \gamma_2 = \gamma_3 = 1$ in the optimization criteria for the design of ψ. Two parameter sets are considered which differ in the role of λ which results in a difference in only the values for λ, χ, and σ. Both are parameter sets such that the forward and inverse problems are matched except for q. The first set of parameters is matched to the forward problem with the exception of $q = 1.0$ (as in the exact inverse problem) and has in addition $\chi = 1.0$, $\lambda = 1.0$, and $\beta = 1.0$. The second set of parameters uses λ instead of σ^2 to control the accuracy of the observations. Then χ is used to compensate in H^{apriori} for the value of $\lambda \neq 1$. Specifically, the quantities $\frac{\lambda}{\sigma^2}$ and $\chi\lambda$ have the same value in both parameter sets but in the first set $\lambda = 1$ while in the second set $\sigma = 1$. For Problem 2 asymptotics both the first and second set of parameters are used. As expected, both calculations yield the same results. This fact amounts to a numerical check on the equations[1] and software. For Problem 1 asymptotics only the first set of parameters are used.

Statistics are collected on $N = 1000$ trials for each value of σ. The underlying fields $\{\phi\}$ are drawn from the probability mass function $\frac{1}{Z}\exp(-\beta H^{\text{apriori}})$ where $Z = \sum_{\{\phi\}}\exp(-\beta H^{\text{apriori}})$ using the Metropolis algorithm[2]. At a given σ, the same set of stochastic realizations of the forward problem is used for all three nontrivial estimators. Since $1000 \approx 2^{10}$ and there are only 2^{17} configurations of the underlying lattice random variables, the lattice variable contribution to the total randomness in the measurements is fairly well sampled. (There are actually fewer distinguishable configurations because translations and inversions are unobservable).

Two measures of performance are considered. Both measures are expectations which are

---

[1] To see that the equations are invariant under parameter changes such that $\frac{\lambda}{\sigma^2}$ and $\chi\lambda$ are constant, it is helpful to introduce a new Lagrange variable τ' to replace τ (see [1, Section 10]) where $\tau' = \tau\lambda$.

[2] The initial condition for the Metropolis algorithm is independent pseudo random binary numbers at each site with equal probability of 0 and 1, the system is equilibrated for 200000 trials, and then the result of every 10000th trial is used.

approximately computed by averaging the results of the N trials[3]. Let $\hat{\phi}_n$ be an estimate of ϕ_n. Because the phase of ϕ_n is 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 $\hat{\phi}_n$. Let $||x||_p = (\sum_n |x_n|^p)^{1/p}$. The first 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,[4] i.e., $E(l_2) = E \min ||\phi - \hat{\phi}||_2$. 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 variability of performance around the average performance is also important. Therefore, for average performance measured in the sense of $E(l_2)$, the variability of performance in the sense of the standard deviation $\mathrm{SD}(l_2)$ is also computed, i.e., $\mathrm{SD}(l_2) = \sqrt{\mathrm{Var} \min ||\phi - \hat{\phi}||_2}$.

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, $f_{\mathrm{perfect}} = \Pr(\mu(\phi, \hat{\phi}) = 0) = E \delta_{\mu(\phi, \hat{\phi}), 0}$ where $\delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$.

Statistics on the trivial estimator are computed in the following fashion. For the previous computations, 13 sets of 1000 random samples from the a priori distribution ($q = 0$) were computed. For this computation take all unordered pairs of not equal sets (78 total), label one as truth and one as estimate (the labeling does not matter since the cost functions are invariant under exchange of truth and estimate), compute performance statistics for that pair, and finally average the performance statistics over all pairs considered. The results are

---

[3]Weighting by the probability mass function is not necessary since the configurations are drawn from the probability mass function.

[4]Recall from [1] that the conditional mean estimator is optimal for the cost function $||\phi - \hat{\phi}||_2^2$. That is, among all estimators $\hat{\phi}(\{y\})$, the conditional mean estimator is the estimator that minimizes $E(||\phi - \hat{\phi}||_2^2 | \{y\})$.

$n_{\text{perfect}} = 74.3077$, $E(l_2) = 1.4792$, and $\text{SD}(l_2) = 0.5346$.

Figures 5 and 6 show plots of $n_{\text{perfect}} = f_{\text{perfect}}N$ (Figure 5) and $E(l_2)$ (Figure 6) versus σ. On both plots the three σ-dependent traces plotted with solid lines show performance for the exact conditional mean estimator (the best performance), the Problem 2 asymptotics approximate estimator, and the Problem 1 asymptotics approximate estimator while the σ-independent trace shows performance for the trivial estimator that disregards the data. The traces for the approximate estimators only differ significantly at high σ where the Problem 2 estimator consistently outperforms the Problem 1 estimator. On Figure 6 the traces plotted with dotted lines show plus/minus one sample standard deviation (i.e., $\text{SD}(l_2)$) relative to $E(l_2)$ for the three nontrivial estimators. The reason that the plots do not have the two highest σ data points for the exact computation is that the number of probabilistically important configurations increases as σ increases. The running time of the current algorithm for the exact computation is sensitive to this number and becomes prohibitively long.

The maximal signal possible at any k in this problem is $L = 17$. Therefore, ignoring the constraints placed on the lattice configuration by the a priori distribution, the large σ region reaches into the realm of rather noisy data. This is born out by the fact that the exact and trivial (observation independent) estimators provide rather similar performance by $\sigma = 6$. Similarly, the small σ region represents rather clean data. The general character of the performance is that at low and high σ all three nontrivial estimators provide rather similar performance while in the transition zone the exact estimator provides good performance to slightly higher σ than the approximate estimators. These general characteristics are common in exact versus approximate nonlinear filters, e.g., in phase-locked loops [4, Chapter 6]. Note that at high σ with respect to the n_{perfect} performance measure the exact estimator performs less well than the approximate estimators, reflecting the fact that the exact conditional mean estimator is not optimal for the n_{perfect} performance criteria.

The fact that the Problem 2 estimator outperforms the Problem 1 estimator at high σ is expected. Recall that the Problem 2 estimator weights the a priori Hamiltonian H^{apriori} by

the asymptotic parameter λ while the Problem 1 estimator does not. Since the estimators are computed in the $\lambda \to \infty$ limit (even though they are used at finite λ), the Problem 2 formulation relies more on a priori information and at high σ a priori information is the only available information.

In Figure 6 the mean plus/minus one standard deviation lines are quite distant from the mean. One of the reasons for this is the discrete nature of the $\hat{\phi}$ estimates. Since $\min ||\phi - \hat{\phi}||_2^2 = \mu(\phi, \hat{\phi})$ and μ takes on only small integer values, the standard deviation of $\min ||\phi - \hat{\phi}||_2$ is larger than might be expected because all deviations from mean performance must be at least a certain minimal distance away. In the Table a complete set of histograms as σ varies are shown for the number of trials (out of N total trials) in which $\mu(\phi, \hat{\phi})$ takes each of its observed values when $\hat{\phi}$ is computed by the exact conditional mean estimator.

At high σ, all three nontrivial estimators give estimates with higher values of $E(l_2)$ than the trivial estimator that disregards the data. I attribute this to the $q = 1$ used to break the symmetries in the nontrivial estimators, which is not equal to the actual value of $q = 0$ used both to select the true configurations from the a priori distribution and to select the estimated configurations in the trivial estimator. One could compute a second trivial estimator which again disregards the data but draws from a distribution with $q = 1$. I did not do this because this estimate is (slightly) more costly to compute than the $q = 0$ estimate and will have poorer performance so there is no logical reason to use such an estimator. In addition there may be a statistical effect since the trivial estimator's performance is based on a far larger number of trials.

In Figure 7 a small subset of the truth configurations used in these simulations are shown. While the small size and the one-dimensional nature of the lattice preclude a really "molecular" appearance, still the basic idea is present–the 1's, which correspond to the atoms, are generally neither adjacent nor isolated in both directions by great distances. Rather, the 1's occur on average one "covalent bond length" from each other. Note that the boundary conditions for this space group are the standard toroidal boundary conditions so that 1's

near an end may be bound to 1's near the other end.

7 A Two-Dimensional Example

In one-dimensional examples there is no concept of bond angle and therefore there are surprisingly few degrees of freedom in a configuration that has a low value for H^{apriori}. In this section a two dimensional example is presented which demonstrates that the approximate estimators can work in situations with these additional degrees of freedom which are only indirectly constrained by the bond-length oriented a priori Hamiltonian. Also, this example, though still small, has many more lattice sites than any of the previous examples.

The lattice measures 9 by 13 sites. The true configuration has 7 occupied sites located at positions (5,10), (2,8), (3,5), (5,3), (4,0), (1,0), and (0,3). The observation noise has variance 0.75 constant over k. The H^{apriori} of Section 2 has parameters $p_1 = 1.0$, $p_2 = -2.0$, $l_1 = 2.3$, and $l_2 = 4.2$. The approximate conditional mean inverse problem uses Problem 2 asymptotics and has parameters $q = 1$, $\chi = 1$, $\lambda = 1$, and $\beta = 1$. The initial condition for the ψ optimization has parameters $\alpha_1 = \frac{1}{9} = 0.111111$ and $\alpha_2 = \frac{1}{13} = 0.0769231$. This choice of α_1 and α_2 balances the linear and quadratic terms in the ψ initial condition in the sense that the linear term $(L - n)$ and the quadratic term $\alpha(L - n)^2$ evaluated at $n = 0$ (see eqn. 2) have the same value in both coordinate directions.

Results for the approximate conditional mean estimator are shown in Figures 8–11. Figure 8 shows the initial condition for ψ as a smooth surface in three dimensions above the two-dimensional lattice array. Figure 9 shows the optimized ψ in the same format as Figure 8. Note, as expected in light of the simple optimization scheme, that the initial and terminal ψ are qualitatively quite similar. Therefore the optimization, which does not explicitly include a measure of symmetry-breaking performance, has probably not greatly altered the symmetry breaking built into the initial condition ψ. Figure 10 shows the true ϕ field and the estimated ϕ field. Since these are binary fields, they are shown as patterns of occupied

sites rather than as projections of surfaces. The estimate is exactly correct, though shifted from the true configuration by an inversion through the origin followed by a translation by the vector (5,11). Better performance cannot be hoped for since inversion and translation do not effect the magnitude of the Fourier transform coefficients. Finally, Figure 11 shows the final (i.e., computed with the optimized ψ) approximation to $E(\phi_n|\{y\})$, the conditional expectation of the lattice field, in the same format as Figures 8 and 9. Figure 10, the estimated ϕ field, results from thresholding Figure 11 at value $\frac{1}{2}$. In light of the large local maxima, it is straightforward to anticipate the result of the thresholding operation.

8   Discussion and Future Directions

In this paper a parameterization of H^{apriori} that is appropriate for crystallography is proposed; data-adaptive ideas for the choice of ψ, the convolutional kernel in the symmetry-breaking $H^{\text{s.b.}}$, are described; and several numerical examples in one and two dimensions on simulated data are given. The examples included a tiny one-dimensional problem where it is computationally practical to compute the estimator performance statistics versus observation noise intensity for the exact conditional mean estimator by brute force and compare with the approximate estimators. These examples, though small in size, demonstrate that these approximations based on the spherical model and small-noise asymptotics provide reasonable performance relative to the exact conditional mean estimator, which is extremely expensive to compute. In the future superior optimization schemes are needed, as discussed in Section 4, in order to allow larger problems to be considered and more general space group symmetries need to be included here and in [1].

9　Acknowledgements

10　Appendix

This appendix is based on the results of [1, Section 10] and uses the definitions of that paper. In this appendix an algorithm is described which efficiently solves for the critical point, that is the constrained global minimum, of βH_λ. The location of the critical point is needed in order to compute the small-noise asymptotic evaluation of the integrals required for the conditional-mean Bayesian estimator. In the process of computing the global minimum, all local minimum are also located.

In the case $\tau = \frac{L}{2} b_{\bar{k},2b}$ for some $\bar{k} \in \{1, \ldots, \frac{L-1}{2}\} - B$ I simply set $\rho_{\bar{k}} = 0$ while in theory $\rho_{\bar{k}}$ is undetermined (though nonnegative) and I should eventually integrate over a manifold when computing the asymptotics of the integrals. The manifold will not be all of $\rho_{\bar{k}} \geq 0$ because if $\rho_{\bar{k}}$ is too large, then the constraint equation will not have a real root for ρ_0.

There are three straightforward initialization steps followed by a loop over at most $\frac{L-1}{2} + 1$ elements

Initialization (1). Define

$$c_k = \frac{L}{2}(b_{k,2a} + b_{k,2b}) \quad k \in \{1, \ldots, \frac{L-1}{2}\}.$$

Initialization (2). Define $k_j \in \{1, \ldots, \frac{L-1}{2}\}$ $j \in \{1, \ldots, \frac{L-1}{2}\}$ a permutation of $\{1, \ldots, \frac{L-1}{2}\}$ such that $c_{k_1} \leq c_{k_2} \leq \ldots \leq c_{k_{\frac{L-1}{2}}}$. Define $j_{\max} = \max\{j | k_j \in \{1, \ldots, \frac{L-1}{2}\} - B\}$, that is, $c_{k_{j_{\max}}}$ is the largest c for which no observation was taken among frequencies $k \neq 0$. If $\{j | k_j \in \{1, \ldots, \frac{L-1}{2}\} - B\} = \emptyset$, then set $j_{\max} = 0$.

Initialization (3). Define $I_j \subset R$ $j \in \{1, \ldots, \frac{L-1}{2} + 1\}$ by

$$I_j = \begin{cases} (-\infty, c_{k_1}) & \text{if } j = 1 \\ [c_{k_{j-1}}, c_{k_j}) & \text{if } j \in \{2, \ldots, \frac{L-1}{2}\} \\ [c_{k_{\frac{L-1}{2}}}, \infty) & \text{if } j = \frac{L-1}{2} + 1 \end{cases}.$$

Loop (1). Fix $j^* \in \{j_{\max} + 1, \ldots, \frac{L-1}{2} + 1\}$. The algorithm will eventually consider every element of this set.

Loop (2). Assume $\tau \in I_{j^*}$.

Loop (3). By Loop (2), if $k_j \in B$ then

$$\rho_{k_j}(\tau) = \begin{cases} \sqrt{\dfrac{(b_{k_j,2a} + b_{k_j,2b}) - \tau \frac{2}{L}}{2b_{k_j,4a}}} & \text{if } j \geq j^* \\ 0 & \text{if } j < j^* \end{cases}$$

while if $k_j \in \{1, \ldots, \frac{L-1}{2}\} - B$ then

$$\rho_{k_j} = 0.$$

Note that these are not yet computable because τ is not yet known.

Loop (4). Use the constraint condition and the gradient condition at $k = 0$ to solve for ρ_0 and τ. Specifically, define

$$d_0(j^*, B) = \frac{2}{L} \sum_{j \in \{j^*, \ldots, \frac{L-1}{2}\} - B} \frac{b_{k_j,2a} + b_{k_j,2b}}{2b_{k_j,4a}}$$

$$d_1(j^*, B) = \frac{2}{L^2} \sum_{j \in \{j^*, \ldots, \frac{L-1}{2}\} - B} \frac{1}{b_{k_j,4a}}$$

where if $j^* = \frac{L-1}{2} + 1$ then $\{j^*, \ldots, \frac{L-1}{2}\} = \emptyset$ and the sums are zero. Both sums are computable at this point. The constraint condition is

$$\frac{1}{L}\rho_0^2 - \rho_0 + d_0(j^*, B) - \tau d_1(j^*, B) = 0 \qquad (3)$$

and the $k = 0$ component of the gradient condition is

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

Assume $d_1(j^*, B) = 0$. Then eqn. 3 simplifies to

$$\frac{1}{L}\rho_0^2 - \rho_0 + d_0(j^*, B) = 0$$

which determines the ρ_0 roots and there are exactly two possibly equal choices. If the roots are complex conjugates then there are no feasible solutions. For each real root, if the root is $\frac{L}{2}$, then τ for that root is undetermined. Otherwise eqn. 4 specifies that

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

Assume $d_1(j^*, B) \neq 0$. Then eqn. 3 can be solved for τ giving

$$\tau = \frac{\frac{1}{L}\rho_0^2 - \rho_0 + d_0(j^*, B)}{d_1(j^*, B)}. \qquad (5)$$

Use of this result in eqn. 4 gives

$$\rho_0^3\left(\frac{2}{L^2} + 4b_{0,4a}d_1(j^*, B)\right) + \rho_0^2\left(\frac{-3}{L}\right)$$

$$+\rho_0\left(-2(b_{0,2a} + b_{0,2b})d_1(j^*, B) + \frac{2}{L}d_0(j^*, B) + 1\right)$$

$$+(-b_{0,1b}d_1(j^*, B) - d_0(j^*, B)) = 0$$

which determines the ρ_0 roots and there are exactly three possibly equal choices. One choice is sure to be real while a pair may be complex conjugates and thus not acceptable. Then, for each real root, eqn. 5 specifies the corresponding τ.

If $\rho_0 \notin R$ or $\tau \notin I_{j^*}$ then discard that solution. For each retained solution $\rho_0 \ \tau$. use the equations of Loop (3) to compute $\rho_k \ k \in \{1, \ldots, \frac{L-1}{2}\}$ and then compute βH_λ.

Loop (5). Among the values of βH_λ found in Step (7) (there may be none) and the minimum value of βH_λ found so far from other j^*, save the minimum value and the τ, ρ_0, ρ_k $k \in \{1, \ldots, \frac{L-1}{2}\}$ for which the value occurred.

Loop (6). If all $j^* \in \{j_{\max} + 1, \ldots, \frac{L-1}{2} + 1\}$ have been considered then the result from the most recent iteration of Loop (5) is optimal and the algorithm terminates. Otherwise, return to Loop (1) and select a new value for j^*.

References

[1] Peter C. Doerschuk. Bayesian signal reconstruction, Markov random fields, and x-ray crystallography. *Journal of the Optical Society of America A*, 1991. To appear.

[2] 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.

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

[4] Andrew J. Viterbi. *Principles of Coherent Communication.* McGraw-Hill Book Company, New York, 1966.
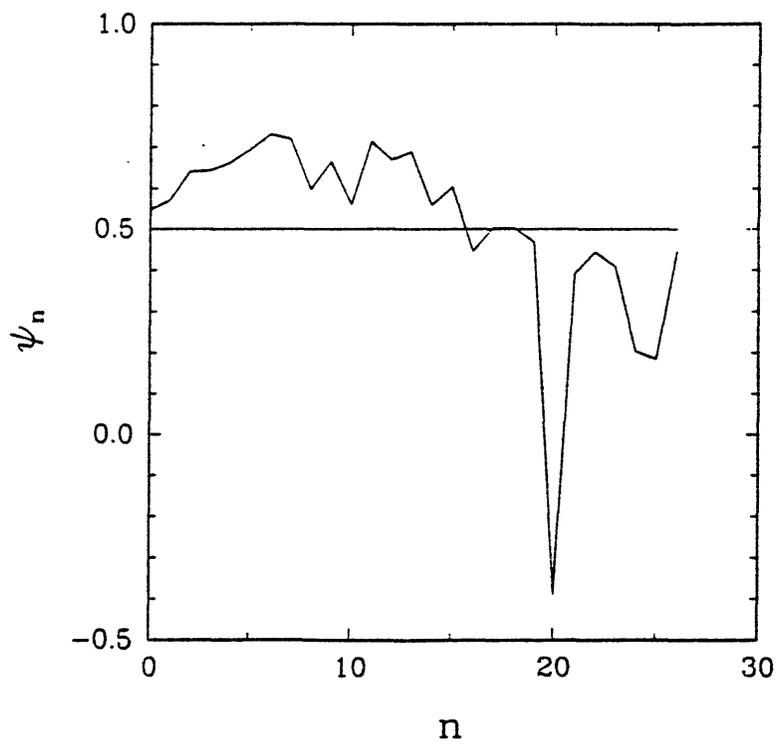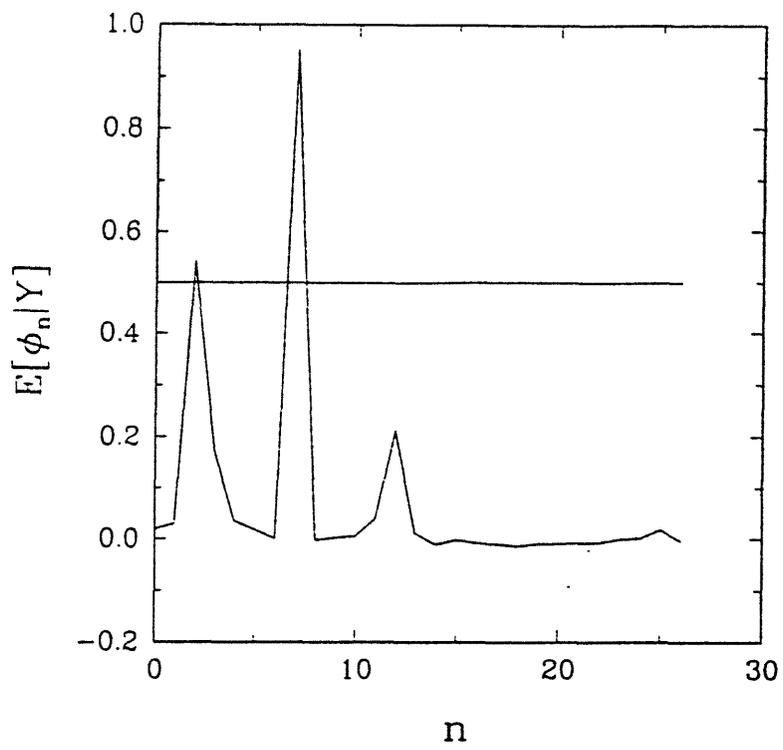
Figure 1: Results for Two Occupied Sites Using Cost C^0. Part a: Conditional Expectations Before Thresholding. Part b: Optimized ψ_n.
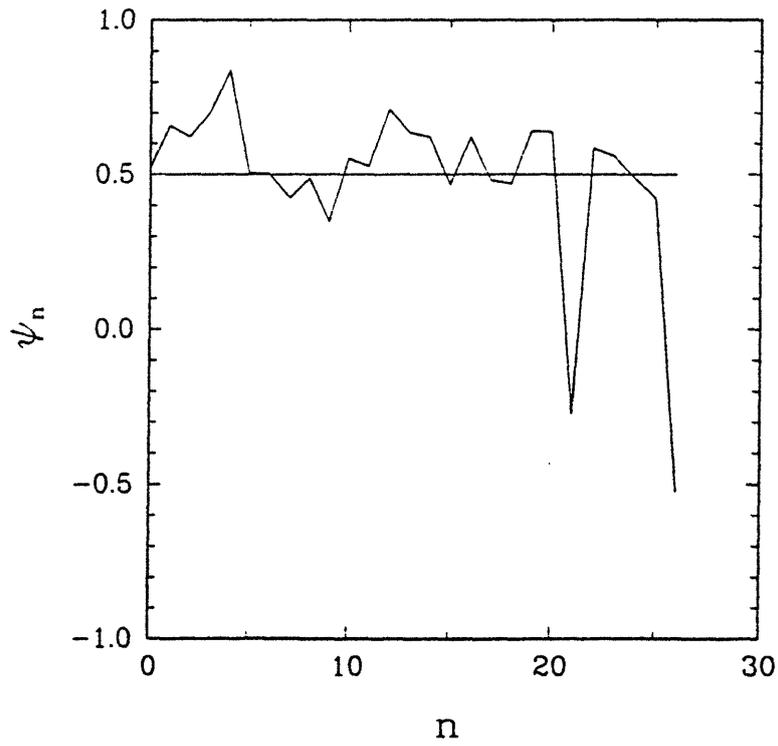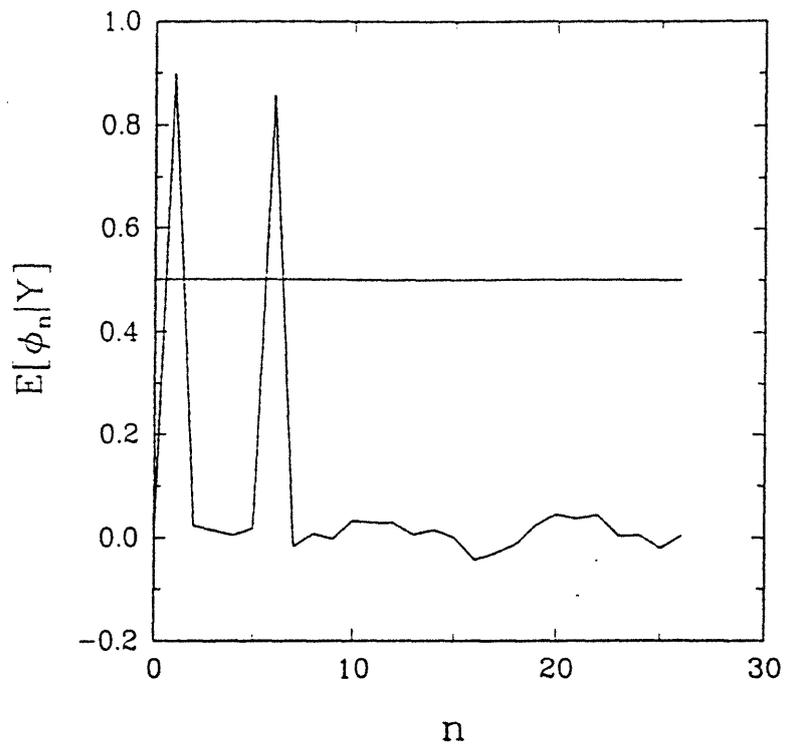
Figure 2: Results for Two Occupied Sites Using Cost C^1. Part a: Conditional Expectations Before Thresholding. Part b: Optimized ψ_n.
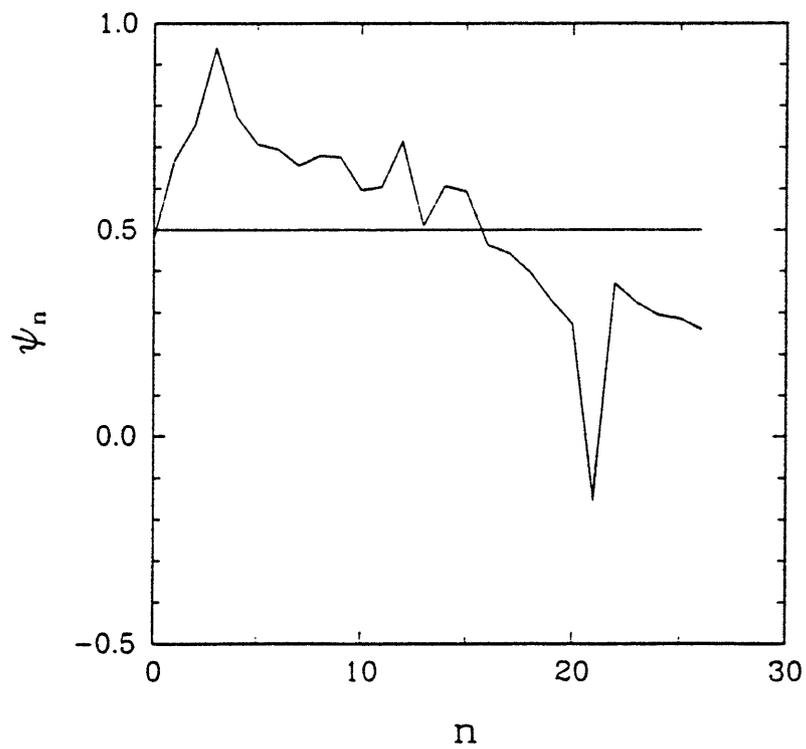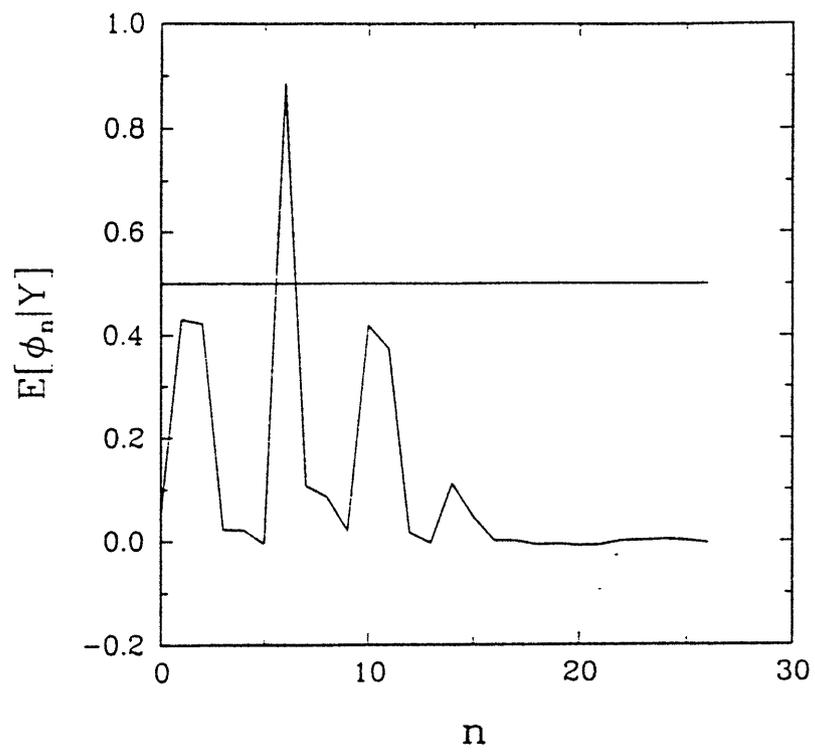
Figure 3: Results for Three Occupied Sites Using Cost C^0. Part a: Conditional Expectations Before Thresholding. Part b: Optimized $\dot{\psi}_n$.
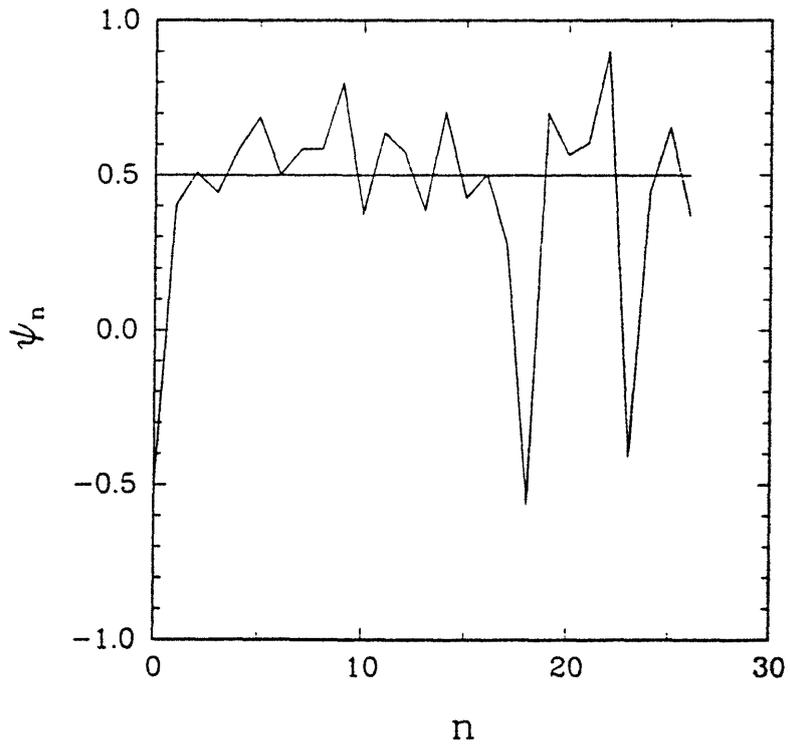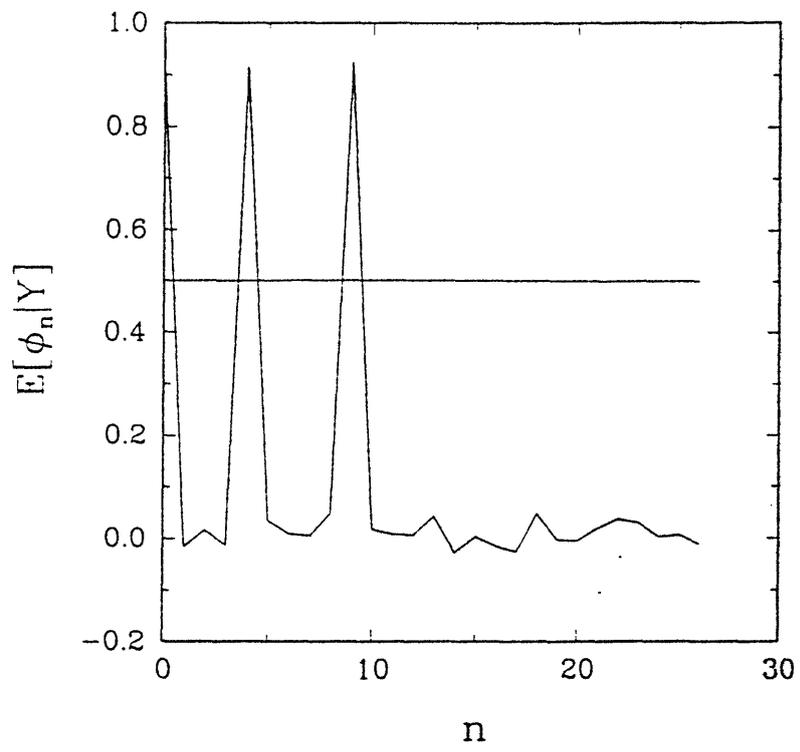
Figure 4: Results for Three Occupied Sites Using Cost C^1. Part a: Conditional Expectations Before Thresholding. Part b: Optimized ψ_n.
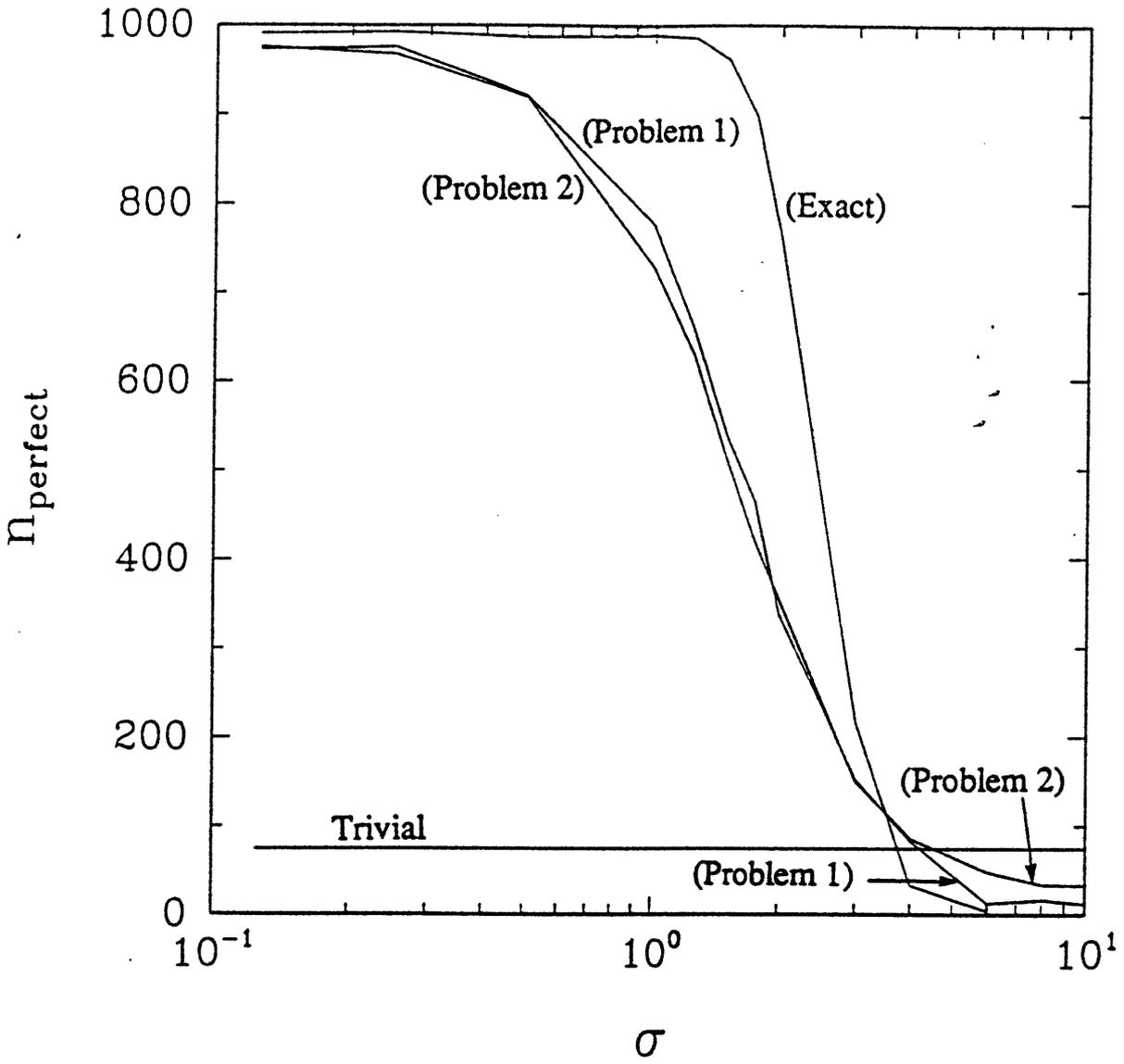
Figure 5: Estimator Performance Statistics: $n_{perfect}$ versus σ for exact, approximate (Problem 1 and Problem 2 asymptotics), and trivial estimators. Recall that $n_{perfect}$ is the number of trials (out of $N = 1000$ total) where (after translation and inversion if required) the reconstruction $\hat{\phi}$ exactly matches the true ϕ and σ is the observation noise standard deviation.
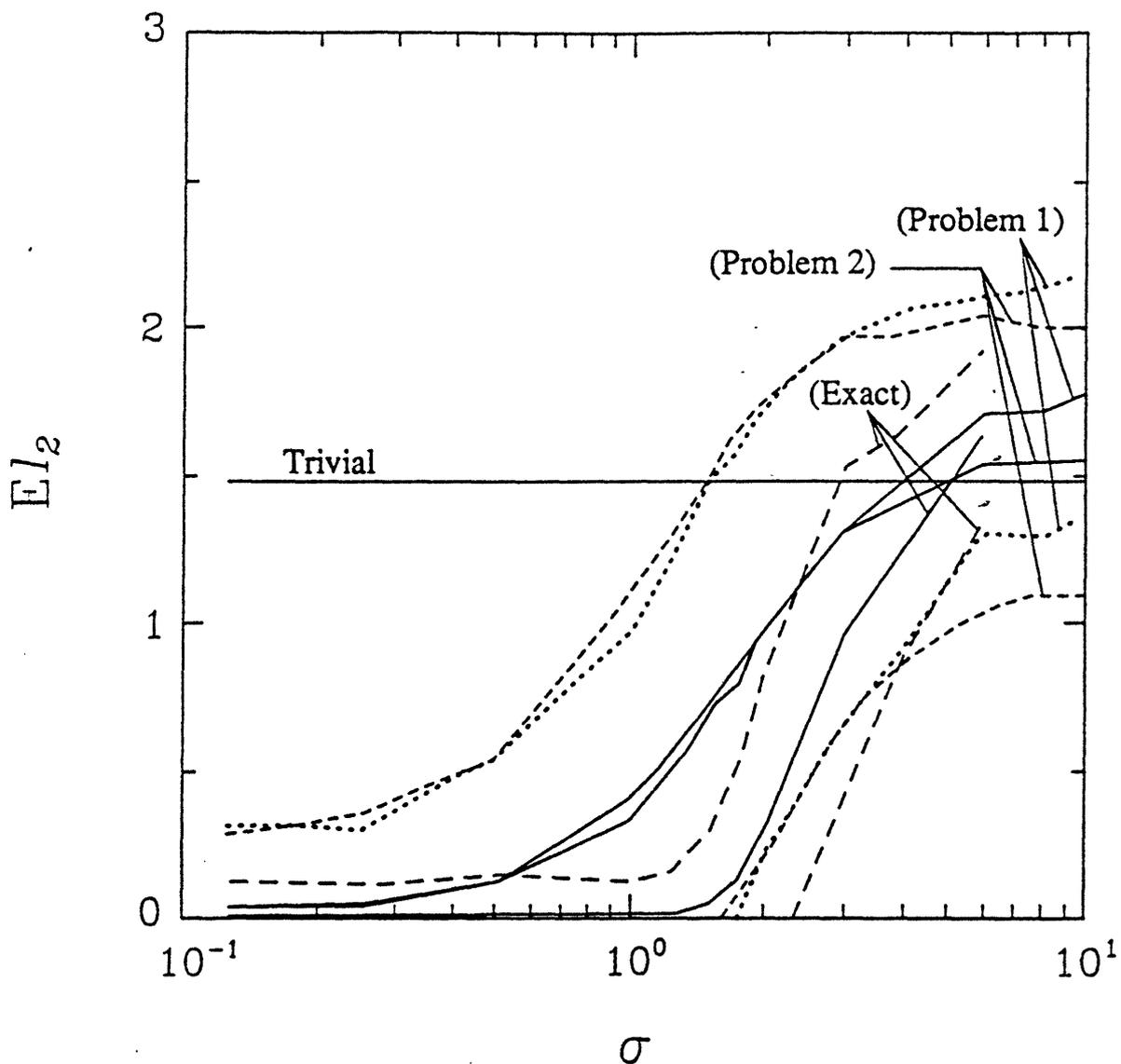
Figure 6: Estimator Performance Statistics: $E(l_2)$ versus σ for exact, approximate (Problem 1 and Problem 2 asymptotics), and trivial estimators. Recall that $E(l_2)$ is the expectation of the l_2 norm of the difference between the reconstruction $\hat{\phi}$ and the true ϕ and σ is the observation noise standard deviation. The traces with labels in parentheses are plus or minus one sample standard deviation relative to the sample mean which carries the same label but without parentheses. The minus one standard deviation traces are truncated (near $\sigma = 2$.) when they become negative.

31

configuration	H^{apriori}	configuration	H^{apriori}
01001001100100100	-4	00010010001000100	-3
00100010010010000	-3	10001001001000100	-5
10010000000000100	-2	00100110010001001	-4
00100100100000000	-2	10010001000100100	-5
10010010001001100	-4	10010010000100100	-4
01000000100100100	-3	00100000000010001	-2
01000100010010010	-5	00100100110010001	-4
01000100100010010	-5	00001000000001001	-1
00010010010010010	-4	10010010001100100	-3
01001000100010010	-5	10001001001001000	-5
00010010010000010	-2	00100100010001001	-5
01000100010000000	-2	00110010010001001	-4
01000100110010000	-2	00010010010010001	-5
00100100100001001	-4	01001001100100100	-4
01001001000001001	0	01001100100000010	-2
00000000011001001	0	00100100001001001	-4
00100100100010011	-4	10010011001001000	-4
10010010011000000	-1	00010000100001001	-2
00100010010001001	-5	01001001001000110	-3
00010010010010001	-5	10001000100000001	1

Figure 7: Example One-dimensional Lattice Configurations and the Corresponding Values of the A Priori Hamiltonian.
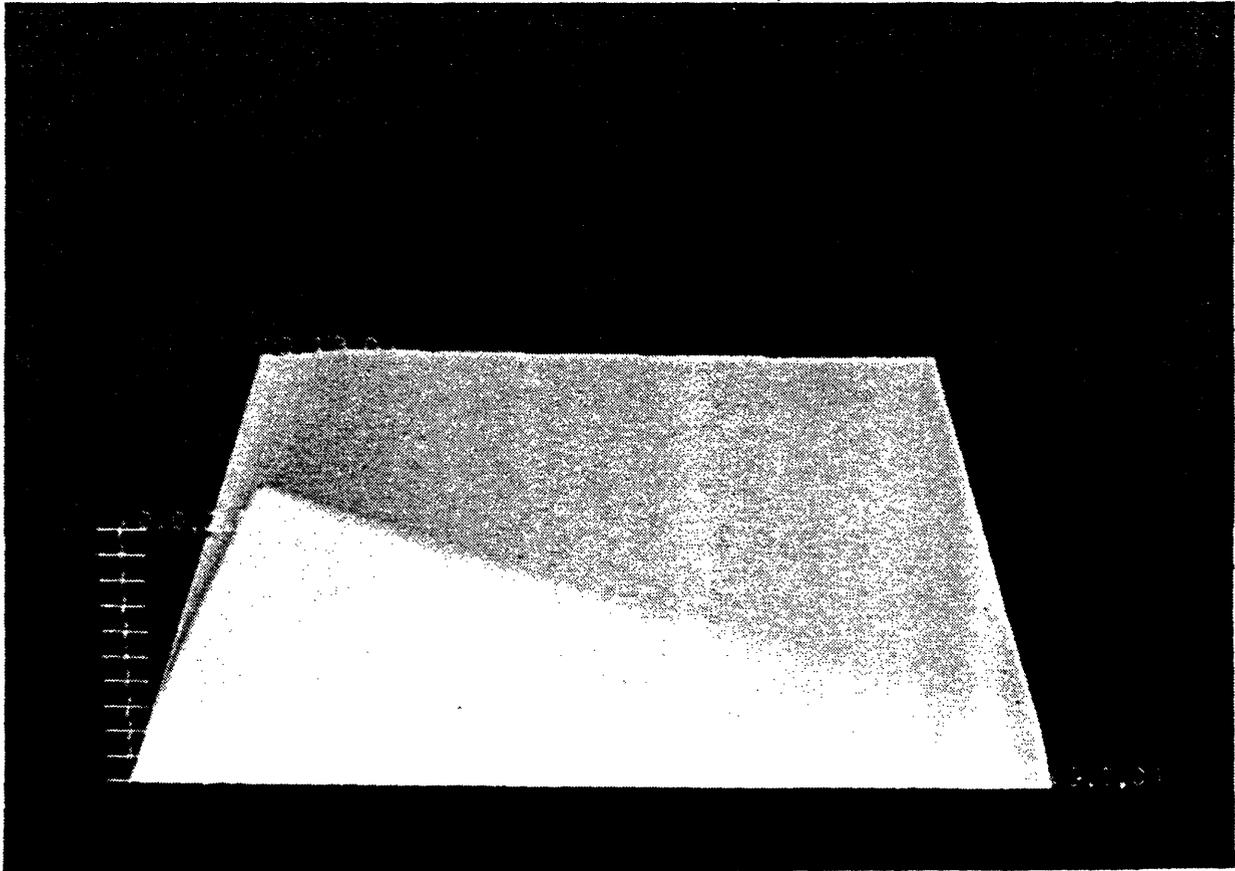
Figure 8: Initial Condition ψ.
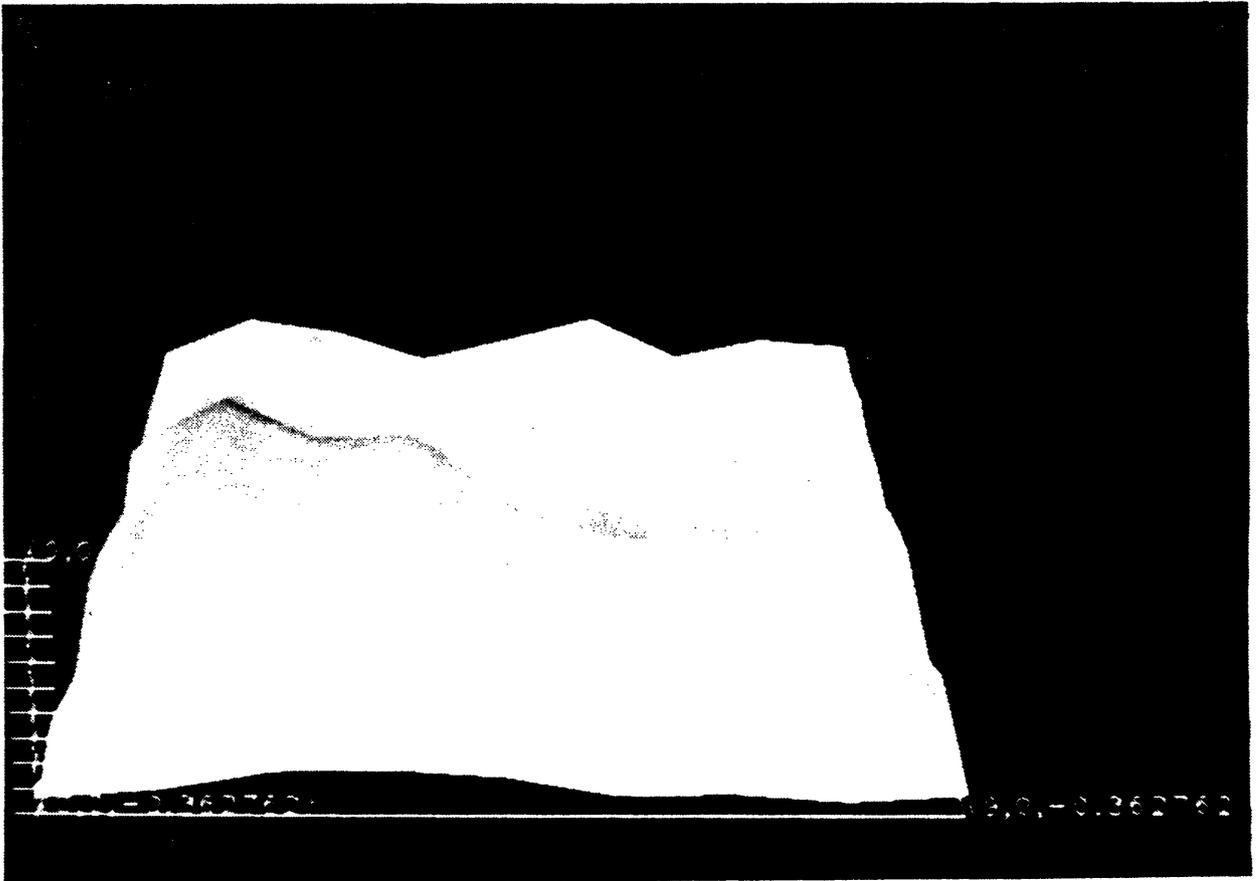
Figure 9: Optimized ψ.
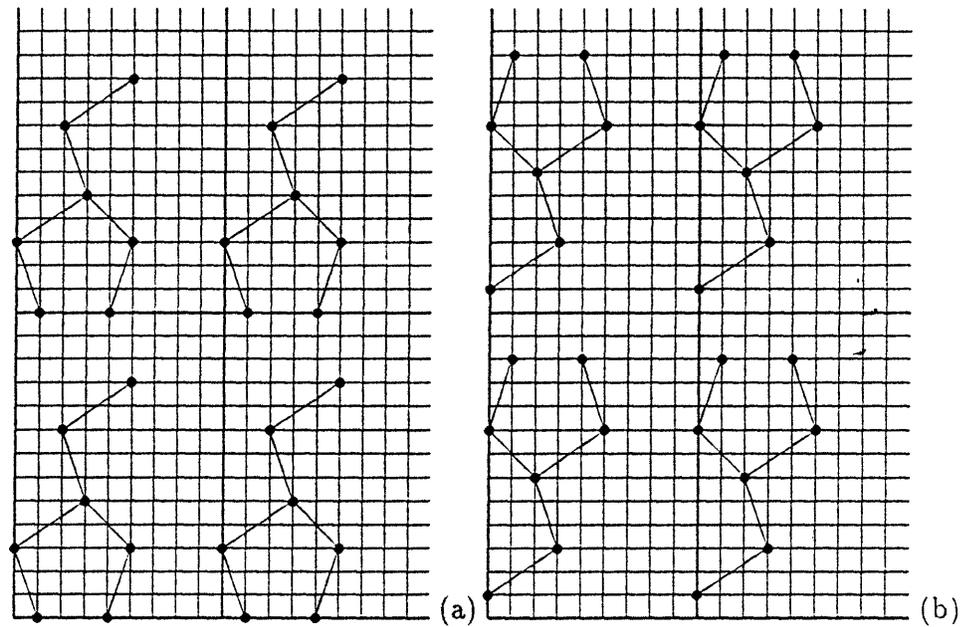
Figure 10: Four Adjacent Copies of the Unit Cell for the 2D Example. Occupied sites (i.e.. n for which $\phi_n = 1$) are denoted by filled circles. All other sites are unoccupied. The boundaries of each unit cell are indicated by thick grid lines. The idealized chemical bond locations are indicated.

Part (a): Original ϕ_n.

Part (b): Reconstructed ϕ_n, which is Figure 11 thresholded at 1/2.

The original and reconstructed ϕ_n agree exactly after an inversion through the origin followed by a translation by the vector (5,11).
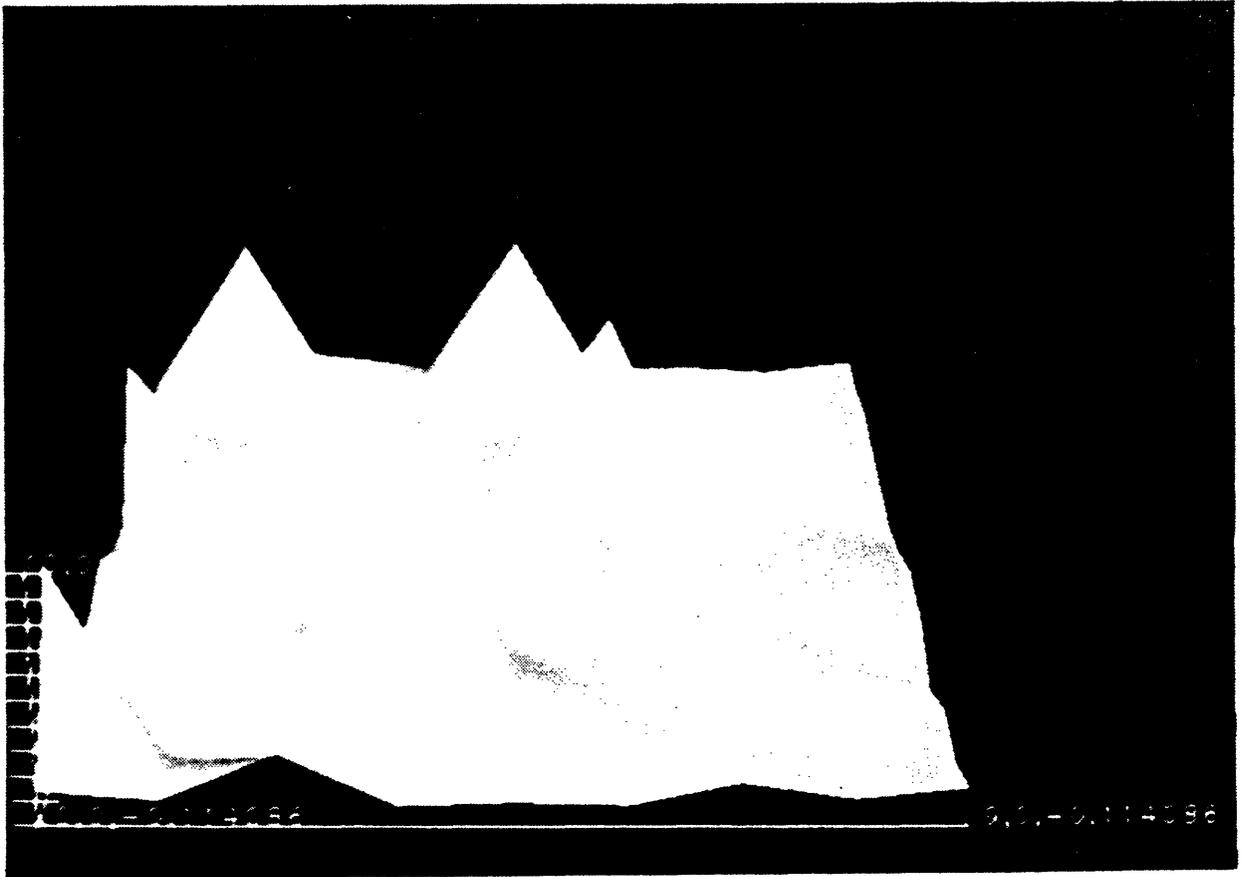
Figure 11: Approximation to $E(\phi_n|\{y\})$.

σ	$\mu(\phi, \hat{\phi})$					
	0	1	2	3	4	5
0.125	991	3	5	1	0	0
0.25	993	5	0	2	0	0
0.5	987	9	3	1	0	0
1.	989	9	2	0	0	0
1.25	986	8	6	0	0	0
1.5	961	22	15	0	1	1
1.75	897	40	58	5	0	0
2.	764	114	113	8	1	0
3.	217	435	241	96	10	1
4.	33	445	308	190	19	5
6.	4	56	312	460	154	14

Table 1: Histograms of $\mu(\phi, \hat{\phi})$ at Each Value of σ for $N = 1000$ Trials of the Exact Conditional Mean Estimator. Each row corresponds to a different value of σ which is listed in the first column. Each column beneath the label $\mu(\phi, \hat{\phi})$ is headed by an integer. The entries in such a column are the number of times $\mu(\phi, \hat{\phi})$ takes that value among the $N = 1000$ trials at each σ.