An Efficient Newton's Method for Entropy Maximization in Phase Determination

Zhijun Wu, George Phillips, Richard Tapia, and Yin Zhang

CRPC-TR99792
May 1999

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

An Efficient Newton's Method for Entropy Maximization in Phase Determination[*]

Zhijun Wu[†], George Phillips[‡], Richard Tapia[§], and Yin Zhang[¶]

Abstract. The joint probability distribution is required for every basis set of structure factors in the Bayesian statistical approach to phase determination. It is computed by maximizing the entropy of the crystal system subject to certain constraints on the structure factors. We propose a Newton's method for the entropy maximization problem. In particular, we show that for m structure factors the method requires only $\mathcal{O}(m \log m)$ floating point operations when the problem structure is exploited. The background of the Bayesian statistical approach to phase determination is introduced. The entropy maximization problem and related solution methods are described. The proposed Newton's method is presented along with its convergence and complexity properties.

Key Words. X-ray crystallography, protein structure determination, entropy maximization, nonlinear optimization, rank-one update, fast Fourier transform

1 Introduction

Entropy maximization is one of the major computational components in the Bayesian statistical approach to phase determination [2, 3, 5, 6, 28]. The problem is formulated to maximize the entropy of a given crystal system subject to a set of structural constraints. The maximum entropy obtained by solving the problem is used to compute the probability distribution of the system with given structure factors.

The entropy of a given crystal system is defined by the following integral,

$$\mathcal{H}(\rho) = - \int_{\mathbf{V}} \rho(r) \log[\rho(r)] dr, \tag{1}$$

where $\mathbf{V}$ is the unit cell and ρ the electron density distribution. Note that without any constraints $\mathcal{H}$ is maximized when ρ is a uniform distribution. Let $\bar{m}$ denote the uniform distribution, $\bar{m} = 1/V$, where V is the volume of the unit cell. Then $\mathcal{H}_{max} = \mathcal{H}(\bar{m}) = \log V$.

The relative entropy of a system ρ with respect to the uniform distribution $\bar{m}$ is defined as the entropy loss of ρ from $\bar{m}$,

$$\mathcal{S}_{\bar{m}}(\rho) = \mathcal{H}(\rho) - \mathcal{H}(\bar{m}) = - \int_{\mathbf{V}} \rho(r) \log[\rho(r)/\bar{m}(r)] dr. \tag{2}$$

Let N be the number of electrons in the system and F^* a given set of structure factors, $F^* = \{F^*_{H_j} : j = 1, \ldots, m\}$, where H_j are three-dimensional integer vectors in the reciprocal space of the crystal lattice. Then the probability distribution $P(F^*)$ of ρ with given structure factors in F^* can be computed by the following formula,

$$P(F^*) = \exp(N\mathcal{S}^*_{\bar{m}}), \tag{3}$$

where $\mathcal{S}^*_{\bar{m}}$ is the optimal value of the entropy maximization problem,

$$\max_{\rho} \quad \mathcal{S}_{\bar{m}}(\rho) \tag{4}$$

$$s.t. \quad \int_{\mathbf{V}} \rho(r) \exp(2\pi i H_j^T r) dr = F^*_{H_j}, \quad j = 1, \ldots, m \tag{5}$$

$$\int_{\mathbf{V}} \rho(r) dr = 1. \tag{6}$$

Note that the objective function in (4) is concave and the constraints are linear. Therefore, the problem is a convex program. In fact, the objective

2

function is even strictly concave, and hence, the solution to the problem must also be unique.

Bricogne [2] reduces the problem (4) to a system of equations, which is equivalent to the dual problem of (4) [1]. A standard Newton's method can be used to solve the equations. While the Newton's method converges fast (local quadratic convergence), it requires $\mathcal{O}(m^3)$ floating point operations in every iteration. This can be a computational bottleneck when m is large. For example, in phase refinement, the entropy maximization problem needs to be solved many times with m possibly in order of thousands. An alternative method is proposed in [2] which computes the inverse of the Jacobian matrix approximately with $\mathcal{O}(m \log m)$ floating point operations. However, the convergence rate of the method is slowed down because of the approximation.

In this paper, we propose a method which has the same convergence rate as the standard Newton's method but computes the exact inverse of the Jacobian matrix in order of $m \log m$ floating point operations. The idea of the method is based on the observation that the Jacobian matrix has a special structure with a positive definite matrix plus a rank-one update. Therefore, by applying the Sherman-Morrison-Woodbury theorem [9, 15, 27], the inverse of the Jacobian can be obtained in a special form, which can be computed via fast Fourier transform in $\mathcal{O}(m \log m)$ floating point operations.

In Section 2, we describe the entropy maximization problem in greater detail. We discuss previous approaches in Section 3, and present our algorithm and related convergence and complexity results in Section 4. We conclude the paper in Section 5.

2   Entropy Maximization

In this section, we derive the equations for solving (4) and show that the solution to the equations can also be obtained by solving a dual problem of (4). Most of the results can be found in [21, 22, 1, 2, 5]. However, they were presented informally, and some were not accurate. We describe them more formally and give mathematical proofs for key facts. In particular, we verify the regularity condition for maximum entropy problem, which is necessary for deriving the entropy equations and the dual entropy problem. In [1], an upper bound for the maximum entropy was established by considering the dual entropy problem, however the strong duality condition when the duality

3

gap is equal to zero was not discussed. We show a necessary and sufficient condition for the strong duality condition to hold. We also provide a proof for the positive definiteness of the Jacobian matrix of the entropy equations, which is also the Hessian of the dual entropy problem.

For convenience, we write the entropy maximization problem (4) in the following general form,

$$\max_\rho \quad \mathcal{S}_{\bar{m}}(\rho) \tag{7}$$

$$s.t. \quad \mathcal{C}_j(\rho) = c_j = F_{H_j}^*, \quad j = 1, \dots, m \tag{8}$$

$$\mathcal{C}_0(\rho) = c_0 = 1, \tag{9}$$

where $\mathcal{C}_j$ are linear constraint functionals defined as

$$\mathcal{C}_j(\rho) = \int_\mathbf{V} \rho(r) C_j(r) dr, \quad j = 0, \dots, m, \tag{10}$$

with

$$C_j(r) = \exp(2\pi i H_j^T r), \quad j = 1, \dots, m \tag{11}$$

$$C_0(r) = 1. \tag{12}$$

We now form the Lagrangian function for (7) as follows,

$$\mathcal{L}(\rho, \lambda_0, \dots, \lambda_m) = \mathcal{S}_{\bar{m}}(\rho) + \sum_{j=0}^m \lambda_j [\mathcal{C}_j(\rho) - c_j]. \tag{13}$$

If ρ is a local maximizer of (7), the partial derivative of the Lagrangian function with respect to ρ is necessarily equal to zero. We then obtain

$$-1 - \log[\rho(r)/\bar{m}(r)] + \sum_{j=0}^m \lambda_j C_j(r) = 0. \tag{14}$$

Solve the equation for ρ to obtain

$$\rho(r) = \bar{m}(r) \exp(\lambda_0 - 1) \exp[\sum_{j=1}^m \lambda_j C_j(r)]. \tag{15}$$

Let $\lambda_0 - 1 = -\log Z$. Then,

$$\rho(r) = \frac{\bar{m}(r)}{Z} \exp[\sum_{j=1}^m \lambda_j C_j(r)]. \tag{16}$$

Since ρ satisfies the normalization constraint (9),

$$\mathcal{C}_0(\rho) = \int_{\mathbf{V}} \rho(r) C_0(r) dr = \int_{\mathbf{V}} \rho(r) dr = 1, \tag{17}$$

we then obtain Z as a function of $\lambda_1, \ldots, \lambda_m$,

$$Z(\lambda_1, \ldots, \lambda_m) = \int_{\mathbf{V}} \bar{m}(r) \exp[\sum_{j=1}^{m} \lambda_j C_j(r)] dr. \tag{18}$$

By applying other constraints (8) to ρ, we have

$$\int_{\mathbf{V}} \frac{\bar{m}(r)}{Z(\lambda_1, \ldots, \lambda_m)} \exp[\sum_{l=1}^{m} \lambda_l C_l(r)] C_j(r) dr = c_j, \tag{19}$$

for $j = 1, \ldots, m$. These equations can be used to determine $\lambda_1, \ldots, \lambda_m$, and hence ρ in terms of (16). A compact form of the equations can be written as

$$\partial_j(\log Z)(\lambda_1, \ldots, \lambda_m) = c_j, \quad j = 1, \ldots, m. \tag{20}$$

We state these results more formally in the following propositions.

Proposition 2.1 *Let ρ be a local maximizer of problem (7). Then there exist a set of parameters $\lambda_0, \ldots, \lambda_m$ such that*

$$\mathcal{S}'_{\bar{m}}(\rho) + \sum_{j=0}^{m} \lambda_j \mathcal{C}'_j(\rho) = 0, \tag{21}$$

$$\mathcal{C}_j(\rho) = \int_{\mathbf{V}} \rho(r) C_j(r) dr = c_j, \quad j = 0, \ldots, m. \tag{22}$$

Proof. Given the fact that C_j and hence $\mathcal{C}'_j$ are linear independent, the regularity condition holds at ρ. Then, there must exist parameters $\lambda_0, \ldots, \lambda_m$ such that the first order necessary condition for ρ to be a local maximizer of (7) is satisfied, which implies that (21) and (22) are necessarily true. Moreover, since (7) is a convex program, the conditions are also sufficient. $\square$

Proposition 2.2 *A set of parameters $\lambda_0, \ldots, \lambda_m$ satisfy the equations in (21) and (22) if and only if the parameters $\lambda_1, \ldots, \lambda_m$ solve the equations,*

$$\nabla(\log Z)(\lambda_1, \ldots, \lambda_m) = c, \tag{23}$$

where $c = (c_1, \ldots, c_m)^T$ and Z is defined as in (18).

5

Proof. The proof is as discussed in the beginning of the section and demonstrated through the derivation from (13) to (20). □

Let G be a function and $\langle G \rangle$ the average value of G by a probability distribution ρ,

$$\langle G \rangle = \int_{\mathbf{V}} \rho(r) G(r) dr. \tag{24}$$

Then it is easy to verify that

$$\partial_j (\log Z) = \langle C_j \rangle, \tag{25}$$

$$\partial_{jk}^2 (\log Z) = \langle C_j \overline{C}_k \rangle - \langle C_j \rangle \langle \overline{C}_k \rangle = \langle (C_j - \langle C_j \rangle) \overline{(C_k - \langle C_k \rangle)} \rangle, \tag{26}$$

where $\overline{(C_k - \langle C_k \rangle)}$ is the complex conjugate of $(C_k - \langle C_k \rangle)$. It implies that the Hessian of $\log Z$, or the Jacobian of the entropy equations (23), is a covariance matrix of the deviation of C_j's from their averaged values.

Proposition 2.3 *The Hessian of $\log Z$ is the covariance matrix of the deviation of C_j's from their averaged values by the probability distribution ρ, and*

$$\nabla^2 (\log Z) = \langle (C - \langle C \rangle)(C - \langle C \rangle)^H \rangle, \tag{27}$$

where $C = (C_1, \ldots, C_m)^T$, $(C - \langle C \rangle)^H$ is the complex conjugate of $(C - \langle C \rangle)$, and $\langle \rangle$ is taken component-wise.

Proof. By the definition of Z in (18),

$$\partial_j (\log Z) = \frac{1}{Z} \partial_j Z \tag{28}$$

$$= \int_{\mathbf{V}} \frac{\bar{m}(r)}{Z(\lambda_1 \ldots, \lambda_m)} \exp[\sum_{l=1}^m \lambda_l C_l(r)] C_j(r) dr \tag{29}$$

$$= \int_{\mathbf{V}} \rho(r) C_j(r) dr = \langle C_j \rangle. \tag{30}$$

It follows that

$$\partial_{jk}^2 (\log Z) = \frac{1}{Z} \partial_{jk}^2 Z - \frac{1}{Z^2} \partial_j Z \partial_k Z \tag{31}$$

$$= \langle C_j \overline{C}_k \rangle - \langle C_j \rangle \langle \overline{C}_k \rangle \tag{32}$$

$$= \langle (C_j - \langle C_j \rangle) \overline{(C_k - \langle C_k \rangle)} \rangle. \tag{33}$$

The Hessian of $\log Z$ is then obtained in the form of (27). □

6

Corollary 2.1 *The Hessian of* $\log Z$ *is positive definite.*

Proof. Let $x = (x_1, \ldots, x_m)^T$ be a nonzero vector and x^H the complex conjugate of x.

$$
\begin{aligned}
x^H \nabla^2 (\log Z) x &= x^H \langle (C - \langle C \rangle)(C - \langle C \rangle)^H \rangle x & (34) \\
&= \langle x^H (C - \langle C \rangle)(C - \langle C \rangle)^H x \rangle & (35) \\
&= \langle |x^H (C - \langle C \rangle)|^2 \rangle & (36) \\
&\geq 0. & (37)
\end{aligned}
$$

Assume that the equality holds for some x,

$$
x^H \nabla^2 (\log Z) x = \langle |x^H (C - \langle C \rangle)|^2 \rangle = 0. \tag{38}
$$

We then have

$$
x^H (C - \langle C \rangle) = 0. \tag{39}
$$

Given the fact that $C_j \neq \langle C_j \rangle$ and $C_j - \langle C_j \rangle$ are linear independent of each other, x must be equal to zero, contradicting to the assumption that x be a nonzero vector. Therefore,

$$
x^H \nabla^2 (\log Z) x = \langle |x^H (C - \langle C \rangle)|^2 \rangle > 0, \tag{40}
$$

and $\nabla^2 (\log Z)$ is positive definite. $\square$

We now show that solving the entropy equation (23) is equivalent to solving the dual problem of (7). According to the standard theory of convex programming [15], the dual problem of (7) is a minimization problem for the Lagrangian function subject to a necessary condition that the partial derivative of the Lagrangian function with respect to ρ is equal to zero, that is,

$$
\min_{\rho, \lambda_0, \ldots, \lambda_m} \quad \mathcal{S}_{\bar{m}}(\rho) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho) - c_j] \tag{41}
$$

$$
s.t. \qquad \mathcal{S}'_{\bar{m}}(\rho) + \sum_{j=0}^{m} \lambda_j \mathcal{C}'_j(\rho) = 0. \tag{42}
$$

Use the condition (42) to obtain ρ as in (15). It then follows that

$$\mathcal{S}_{\bar{m}}(\rho) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho) - c_j] \tag{43}$$

$$= -\int_{\mathbf{V}} \rho(r) \log[\rho(r)/\bar{m}(r)] dr + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho) - c_j] \tag{44}$$

$$= -\int_{\mathbf{V}} \rho(r) [\sum_{j=0}^{m} \lambda_j C_j(r) - 1] dr + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho) - c_j] \tag{45}$$

$$= \int_{\mathbf{V}} \rho(r) dr - \sum_{j=0}^{m} \lambda_j c_j \tag{46}$$

$$= \int_{\mathbf{V}} \bar{m}(r) \exp[\sum_{j=0}^{m} \lambda_j C_j(r) - 1] dr - \sum_{j=0}^{m} \lambda_j c_j. \tag{47}$$

Then, problem (41) becomes

$$\min_{\lambda_0,\ldots,\lambda_m} \int_{\mathbf{V}} \bar{m}(r) \exp[\sum_{j=0}^{m} \lambda_j C_j(r) - 1] dr - \sum_{j=0}^{m} \lambda_j c_j. \tag{48}$$

Let the partial derivative of the objective function with respect to λ_0 be equal to zero and solve the equation for λ_0. Then, the problem can further be reduced to

$$\min_{\lambda} \mathcal{D}(\lambda) = \log Z(\lambda) - c^T \lambda, \tag{49}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)^T$, $c = (c_1, \ldots, c_m)^T$, and Z is defined the same as in (18). In general, we have for any dual feasible $\rho, \lambda_0, \ldots, \lambda_m$,

$$\mathcal{S}_{\bar{m}}(\rho) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho) - c_j] = \log Z(\lambda) - c^T \lambda = \mathcal{D}(\lambda). \tag{50}$$

A necessary condition for λ to be a solution to problem (49) is that the gradient of the objective function at λ is equal to zero, and therefore,

$$\nabla(\log Z)(\lambda) = c, \tag{51}$$

which is the same as the equation in (23). Since the Hessian of the objective function (49) is equal to $\nabla^2(\log Z)$ which is positive definite, the necessary condition (51) is also sufficient, and it determines λ uniquely.

As a standard result from the duality theory for convex programming, we obtain the following relationship between problems (7) and (49).

Proposition 2.4 *Let ρ be any feasible solution of problem (7). Then,*

$$\mathcal{S}_{\bar{m}}(\rho) \leq \mathcal{D}(\lambda) \tag{52}$$

for any λ. If ρ is the maximizer of (7) and λ the minimizer of (49), the equality will hold, and vice versa.

Proof. Consider the dual problem (41). Given any $\lambda = (\lambda_1, \ldots, \lambda_m)^T$, let $\lambda_0 = 1 - \log Z(\lambda_1, \ldots, \lambda_m)$, and

$$\rho_\lambda(r) = \frac{\bar{m}(r)}{Z(\lambda_1, \ldots, \lambda_m)} \exp[\sum_{j=1}^{m} \lambda_j C_j(r)]. \tag{53}$$

Then $\lambda_0, \lambda_1, \ldots, \lambda_m$, and ρ_λ together satisfy the dual constraint (42). In other words, they form a feasible solution to the dual problem (41).

Let ρ be a feasible solution to the primal problem (7). Since $\mathcal{S}_{\bar{m}}$ is concave and $\mathcal{C}_j$ are linear, it follows that

$$\begin{aligned}
\mathcal{S}_{\bar{m}}(\rho) &\leq \mathcal{S}_{\bar{m}}(\rho_\lambda) + \mathcal{S}'_{\bar{m}}(\rho_\lambda)(\rho - \rho_\lambda) \tag{54} \\
&= \mathcal{S}_{\bar{m}}(\rho_\lambda) - \sum_{j=0}^{m} \lambda_j \mathcal{C}'_j(\rho_\lambda)(\rho - \rho_\lambda) \tag{55} \\
&\leq \mathcal{S}_{\bar{m}}(\rho_\lambda) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho_\lambda) - \mathcal{C}_j(\rho)] \tag{56} \\
&= \mathcal{S}_{\bar{m}}(\rho_\lambda) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho_\lambda) - c_j]. \tag{57}
\end{aligned}$$

By the definition of ρ_λ and its dual feasibility along with $\lambda_0, \ldots, \lambda_m$,

$$\mathcal{S}_{\bar{m}}(\rho_\lambda) + \sum_{j=0}^{m} \lambda_j [\mathcal{C}_j(\rho_\lambda) - c_j] = \log Z(\lambda) - c^T \lambda = \mathcal{D}(\lambda). \tag{58}$$

The inequality (52) is thus proved.

Suppose that the equality of (52) holds for some ρ^* and λ^*. Since $\mathcal{S}_{\bar{m}}(\rho) \leq \mathcal{D}(\lambda)$ for any primal feasible ρ and any λ, there cannot be such $\rho \neq \rho^*$ or $\lambda \neq \lambda^*$ that

$$\mathcal{S}_{\bar{m}}(\rho) > \mathcal{S}_{\bar{m}}(\rho^*) = \mathcal{D}(\lambda^*) > \mathcal{D}(\lambda). \tag{59}$$

9

Therefore, ρ^* must be optimal for (7) and λ^* for (49).

Now let ρ^* be the maximizer of (7) and $\lambda_0^*, \lambda_1^*, \ldots, \lambda_m^*$ the corresponding Lagrangian multipliers. Then ρ^* and $\lambda_0^*, \lambda_1^*, \ldots, \lambda_m^*$ must also be dual feasible. It follows that

$$
\begin{align}
\mathcal{D}(\lambda^*) &= \log Z(\lambda^*) - c^T \lambda^* \tag{60} \\
&= \mathcal{S}_{\bar{m}}(\rho^*) + \sum_{j=0}^{m} \lambda_j^* [\mathcal{C}_j(\rho^*) - c_j] \tag{61} \\
&= \mathcal{S}_{\bar{m}}(\rho^*). \tag{62}
\end{align}
$$

Since $\mathcal{D}(\lambda) \geq \mathcal{S}_{\bar{m}}(\rho^*) = \mathcal{D}(\lambda^*)$ for all λ, λ^* must be optimal for (49). $\square$

3 Previous Approaches

The problem (49) can be solved by a standard Newton's method, as proposed by Bricogne [2]. The Newton iteration for the problem can be formulated as follows.

$$
\lambda^{(l+1)} = \lambda^{(l)} - \alpha^{(l)} [\nabla^2 (\log Z)(\lambda^{(l)})]^{-1} [\nabla (\log Z)(\lambda^{(l)}) - c], \tag{63}
$$

where $\alpha^{(l)}$ is a step length. Since $\nabla^2 (\log Z)$ is always positive definite, the Newton's direction is descent at any point. With a line search procedure, the method will be able to decrease the function value in every step. If the function is bounded below, which is the case for problem (49), the method will eventually converge to the minimum. Moreover, it converges quadratically when the iterate is close to the optimal solution [9].

We now consider the computation of the Newton step (63). ¿From the discussion in the previous section,

$$
\begin{align}
\partial_j (\log Z) &= \langle C_j \rangle, \tag{64} \\
\nabla (\log Z) &= \langle C \rangle, \tag{65}
\end{align}
$$

and

$$
\begin{align}
\partial_{jk}^2 (\log Z) &= \langle C_j \overline{C}_k \rangle - \langle C_j \rangle \langle \overline{C}_k \rangle \tag{66} \\
\nabla^2 (\log Z) &= \langle C C^H \rangle - \langle C \rangle \langle C^H \rangle. \tag{67}
\end{align}
$$

10

By its definition,

$$\langle C_j \rangle = \int_{\mathbf{V}} \rho(r) C_j(r) dr = F_{H_j}. \tag{68}$$

So, in terms of structure factors,

$$\partial_j (\log Z) = F_{H_j}, \tag{69}$$
$$\partial_{jk}^2 (\log Z) = F_{H_j - H_k} - F_{H_j} F_{-H_k}, \tag{70}$$

and in a general form,

$$\nabla(\log Z) = F, \tag{71}$$
$$\nabla^2(\log Z) = K - F F^H, \tag{72}$$

where $F = (F_{H_1}, \ldots, F_{H_m})^T$, F^H is the complex conjugate of F, and K a matrix with $K_{jk} = F_{H_j - H_k}$.

Given any $\lambda^{(l)}$, we can immediately construct a density distribution function $\rho^{(l)}$ as in (16), and compute $\nabla(\log Z)(\lambda^{(l)})$ and $\nabla^2(\log Z)(\lambda^{(l)})$ with the formulas (69), (70), (71), and (72).

$$\nabla(\log Z)(\lambda^{(l)}) = F^{(l)}, \tag{73}$$
$$\nabla^2(\log Z)(\lambda^{(l)}) = K^{(l)} - F^{(l)} [F^{(l)}]^H, \tag{74}$$

with

$$\partial_j (\log Z)(\lambda^{(l)}) = F_{H_j}^{(l)}, \tag{75}$$
$$\partial_{jk}^2 (\log Z)(\lambda^{(l)}) = F_{H_j - H_k}^{(l)} - F_{H_j}^{(l)} F_{-H_k}^{(l)}, \tag{76}$$

where for any H_j,

$$F_{H_j}^{(l)} = \int_{\mathbf{V}} \rho^{(l)}(r) \exp(2\pi H_j^T r) dr. \tag{77}$$

Since all structure factors $F_{H_j}^{(l)}$ can be computed once in $\mathcal{O}(m \log m)$ calculations with fast Fourier transform, the gradient $\nabla(\log Z)(\lambda^{(l)})$ and the Hessian matrix $\nabla^2(\log Z)(\lambda^{(l)})$ can be assembled in $\mathcal{O}(m \log m)$ computation time. However, the inverse of the Hessian matrix requires matrix factorization which takes $\mathcal{O}(m^3)$ time. Therefore, the time complexity for the whole

11

Newton iteration (63) is $\mathcal{O}(m^3)$. With such an order of complexity, the computation of the Newton step can be a computational bottleneck when m is large. It certainly becomes impractical when the procedure is called repeatedly as in the Bayesian statistical method for structure determination.

An alternative approach to the problem is to compute the inverse of the Hessian matrix approximately, without using matrix factorization, and hence to reduce the total computation time for the Newton iteration. There are many ways to approximate the inverse Hessian such as the BFGS method. Bricogne [2] suggested to use the matrix K in (71) as an approximation to the Hessian $\nabla^2(\log Z)$. This matrix is known as the Karle-Hauptman matrix [23]. While its elements can be obtained by doing a fast Fourier transform for the corresponding density distribution function, the elements of the inverse can also be obtained by the same procedure for the inverse of the density distribution function. Let K^{-1} be the inverse of K. Then $[K^{-1}]_{jk} = E_{H_j - H_k}$, where

$$E_{H_j - H_k} = \int_{\mathbf{V}} \rho^{-1}(r) \exp[2\pi(H_j - H_k)^T r] dr. \qquad (78)$$

Since the inverse of K can be computed by fast Fourier transform for the inverse of the density distribution function, the computation for the iteration (63), when $\nabla^2 \log Z$ is approximated by K, can be arranged to require only $\mathcal{O}(m \log m)$ calculations. However, the fast convergence property of the original Newton's method is no longer guaranteed by the approximation method. As a trade-off between fast convergence and low complexity, an ad hoc approach is taken in the crystallography software BUSTER [5]: The Newton's method is used as default, but switched to the approximation method when more phases are included in the refinement process and the corresponding entropy maximization problem becomes very large.

4   An Efficient Newton's Method

We propose a method which uses the same Newton iteration (63), but computes the inverse of the Hessian differently, requiring only $\mathcal{O}(m \log m)$ calculations. The method is based on the observation that the Hessian $\nabla^2(\log Z)$ consists of the Karle-Hauptman matrix K with a matrix update FF^H. Since K is positive definite, the Sherman-Morrison-Woodbury formula can be used to derive the inverse of $\nabla^2(\log Z)$ as the inverse of K plus a simple matrix

update. The inverse of K can be obtained by doing a fast Fourier transform for the inverse of the density distribution function ρ. Therefore, the inverse of K and hence of $\nabla^2(\log Z)$ can be computed in only $\mathcal{O}(m \log m)$ floating point operations. From algorithmic point of view, this method is still the Newton's method. However, computationally, it requires only order of $m \log m$ computation, and can thereby be applied to large-scale problems. While requiring the same order of computation as the approximation method, this method has the advantage of the Newton's method and converges to an optimal solution quadratically when the iteration is close to the solution. It also provides an accurate Hessian estimate at the solution, which subsequent computation can benefit from.

We first present the method for computing the inverse of the Hessian and then the whole algorithm. Some of the computational issues will be discussed.

The following proposition is a result of extending the Sherman-Morrison-Woodbury Theorem (see, for example, [9]) to positive definite matrices.

Proposition 4.1 *Let T and S be two Hermite matrices, U a vector, and*

$$T = S - UU^H. \tag{79}$$

Let $\sigma = U^H S^{-1} U$. Then T is positive definite if and only if S is positive definite and $\sigma < 1$. Furthermore,

$$T^{-1} = S^{-1} + \frac{S^{-1}UU^H S^{-1}}{1 - \sigma}. \tag{80}$$

Proof. We first show that if T is positive definite, S is positive definite and $\sigma < 1$, and the inverse of T can be computed by (80).

If T is positive definite, $S = T + UU^H$ must be positive definite. Then, T^{-1} and S^{-1} exist and are also positive definite. Multiply (79) by S^{-1} from left and by T^{-1} from right to obtain

$$S^{-1} = T^{-1} - S^{-1}UU^H T^{-1}. \tag{81}$$

It follows that

$$U^H S^{-1} U = (1 - U^H S^{-1} U)U^H T^{-1} U > 0, \tag{82}$$

13

and $U^H S^{-1} U = \sigma$ must be less than zero. From (81), we have

$$T^{-1} = S^{-1} + S^{-1} U U^H T^{-1}. \tag{83}$$

Note that in the second part of the formula, if we substitute T^{-1} recursively,

$$
\begin{aligned}
U U^H T^{-1} &= U U^H (S^{-1} + S^{-1} U U^H T^{-1}) & (84) \\
&= U U^H S^{-1} + U U^H S^{-1} U U^H T^{-1} & (85) \\
&= U U^H S^{-1} + U^H S^{-1} U U U^H T^{-1} & (86) \\
&= U U^H S^{-1} + \sigma U U^H T^{-1}. & (87)
\end{aligned}
$$

Therefore,

$$U U^H T^{-1} = \frac{U U^H S^{-1}}{1 - \sigma}. \tag{88}$$

Substitute (88) to (83) to obtain

$$T^{-1} = S^{-1} + \frac{S^{-1} U U^H S^{-1}}{1 - \sigma}. \tag{89}$$

We now show that if S is positive definite and $\sigma < 1$, T must be positive definite, and the same formula (80) for the inverse of T follows.

If S is positive definite and $\sigma < 1$, we can construct a matrix,

$$T' = S^{-1} + \frac{S^{-1} U U^H S^{-1}}{1 - \sigma}. \tag{90}$$

It is easy to verify that T' is positive definite and

$$T'T = T'(S - U U^H) = (S - U U^H)T' = I. \tag{91}$$

Therefore, T' is the inverse of T and T must be positive definite. $\square$

We now consider the Hessian matrix $\nabla^2(\log Z)$,

$$\nabla^2(\log Z) = K - F F^H, \tag{92}$$

where K is the Karle-Hauptman matrix, $K_{jk} = F_{H_j - H_k}$, and $F = (F_{H_1}, \ldots, F_{H_m})^T$. By applying Proposition 4.1, we obtain the following results.

Proposition 4.2 *Let $\sigma = F^H K^{-1} F$ and $[K^{-1}]_{jk} = E_{H_j - H_k}$. Then,*

$$0 < \sigma = F^H K^{-1} F = \sum_{j,k} \overline{F}_{H_j} E_{H_j - H_k} F_{H_k} < 1. \tag{93}$$

Proof. The result follows directly from Proposition 4.1 and the fact that $\nabla^2 (\log Z)$ is positive definite. $\square$

Proposition 4.3 *Let $\sigma = F^H K^{-1} F$. Then, the inverse of $\nabla^2 (\log Z)$ can be computed by the following formula.*

$$[\nabla^2 (\log Z)]^{-1} = K^{-1} + \frac{K^{-1} F F^H K^{-1}}{1 - \sigma}. \tag{94}$$

Proof. Since $\nabla^2 (\log Z)$ is positive definite, K is also positive definite, and $\sigma < 1$ by Proposition 4.1. The formula can then be derived by using (80) for $\nabla^2 (\log Z)$. $\square$

By using the formula (80), we can compute the Newton step (63) directly without doing numerical factorization, since the inverse of K can be obtained from a Fourier transform of the inverse electron density distribution. Let us write the iteration (63) in the following form,

$$\lambda^{(l+1)} = \lambda^{(l)} - \alpha^{(l)} [\nabla^2 (\log Z)^{(l)}]^{-1} [\nabla (\log Z)^{(l)} - c], \tag{95}$$

where $\nabla^2 (\log Z)^{(l)}$ and $\nabla (\log Z)^{(l)}$ are the Hessian and gradient of $\log Z$ at $\lambda^{(l)}$. Since c is a vector of known structure factors F^*, and

$$\nabla (\log Z)^{(l)} = F^{(l)} \tag{96}$$

$$\nabla^2 (\log Z)^{(l)} = K^{(l)} - F^{(l)} [F^{(l)}]^H, \tag{97}$$

the iteration (95) is equivalent to

$$\lambda^{(l+1)} = \lambda^{(l)} - \alpha^{(l)} [K^{(l)} - F^{(l)} [F^{(l)}]^H]^{-1} [F^{(l)} - F^*]. \tag{98}$$

With the formula (80), we can further write (98) as

$$\lambda^{(l+1)} = \lambda^{(l)} + \alpha^{(l)} \Delta \lambda^{(l)} \tag{99}$$

with

$$\Delta \lambda^{(l)} = v^{(l)} + \frac{[F^{(l)}]^H v^{(l)}}{1 - [F^{(l)}]^H u^{(l)}} u^{(l)}, \tag{100}$$

15

where

$$v^{(l)} = [K^{(l)}]^{-1}[F^* - F^{(l)}], \tag{101}$$

$$u^{(l)} = [K^{(l)}]^{-1}F^{(l)}. \tag{102}$$

Figure 1 shows an outline of the proposed Newton's algorithm with the iterates computed using the above formulas. Note that in step 2b, a Fourier transform is required to obtain all structure factors $F_j^{(l)}$, and the cost is in the order of $m \log m$. In step 2c, $[K^{(l)}]^{-1}$ can be obtained by a Fourier transform which requires another $\mathcal{O}(m \log m)$ time. The remaining work is to form two matrix-vector products, which may take $\mathcal{O}(m^2)$ time if done explicitly. However, the matrix-vector products can be computed by combining the Fourier transform for $[K^{(l)}]^{-1}$ with the vectors, each of which then requiring only $\mathcal{O}(m \log m)$ calculations. So in total, the algorithm requires $\mathcal{O}(m \log m)$ floating point operations or computing time.

To complete the section, we verify the facts that the inverse of K as well as the matrix-vector products $K^{-1}(F^* - F)$ and $K^{-1}F$ can all be obtained through certain forms of Fourier transforms.

Proposition 4.4 *Let K be a Karle-Hauptman matrix, $K_{jk} = F_{H_j - H_k}$, where*

$$F_{H_j - H_k} = \int_{\mathbf{V}} \rho(r) \exp[2\pi i (H_j - H_k)^T r] dr. \tag{103}$$

Then K^{-1} can be obtained with $[K^{-1}]_{jk} = E_{H_j - H_k}$, where

$$E_{H_j - H_k} = \int_{\mathbf{V}} \rho^{-1}(r) \exp[2\pi i (H_j - H_k)^T r] dr. \tag{104}$$

Proof. Let $L = K K^{-1}$. We show that L is an identity matrix. By the definitions of K and K^{-1},

$$L_{jk} = \sum_l F_{H_j - H_l} E_{H_l - H_k} \tag{105}$$

$$= \sum_l F_{H_j - H_l} \int_{\mathbf{V}} \rho^{-1}(r) \exp[2\pi i (H_l - H_k)^T r] dr \tag{106}$$

$$= \int_{\mathbf{V}} \rho^{-1}(r) \exp[2\pi i (H_j - H_k)^T r] \tag{107}$$

$$\sum_l F_{H_j - H_l} \exp[-2\pi i (H_j - H_l)^T r] dr \tag{108}$$

16

The Proposed Newton's Algorithm

1. Input initial $\lambda^{(0)}$. Set $l = 0$.

2. Repeat

   (a) Compute

   $$Z(\lambda^{(l)}) = \int_{\mathbf{V}} \bar{m}(r) \exp[\sum_{j=1}^{m} \lambda_j^{(l)} C_j(r)] dr$$

   $$\rho^{(l)}(r) = \int_{\mathbf{V}} \frac{\bar{m}(r)}{Z(\lambda^{(l)})} \exp[\sum_{j=1}^{m} \lambda_j^{(l)} C_j(r)] dr$$

   (b) Compute, for $j = 1, \ldots, m$,

   $$F_j^{(l)} = \int_{\mathbf{V}} \rho^{(l)}(r) \exp(2\pi i H_j^T r) dr$$

   (c) Set

   $$\delta^{(l)} = F^* - F^{(l)}$$
   $$v^{(l)} = [K^{(l)}]^{-1} \delta^{(l)}$$
   $$u^{(l)} = [K^{(l)}]^{-1} F^{(l)}$$

   (d) Compute

   $$\Delta\lambda^{(l)} = v^{(l)} + \frac{[F^{(l)}]^H v^{(l)}}{1 - [F^{(l)}]^H u^{(l)}} u^{(l)}$$
   $$\lambda^{(l+1)} = \lambda^{(l)} + \alpha^{(l)} \Delta\lambda^{(l)}$$
   $$l = l + 1$$

   (e) If the optimality condition is satisfied, go to 3.

3. Set $\lambda^* = \lambda^{(l)}$, $\rho^* = \rho^{(l)}$. Stop.

Figure 1: Outline of the proposed Newton's algorithm

17

$$= \int_{\mathbf{V}} \rho^{-1}(r) \exp[2\pi i (H_j - H_k)^T r] \rho(r) dr \qquad (109)$$

$$= \int_{\mathbf{V}} \exp[2\pi i (H_j - H_k)^T r] dr. \qquad (110)$$

It is easy to see that $L_{jk} = 0$ if $j \neq k$ and $L_{jk} = 1$ if $j = k$. L is indeed an identity matrix. The same result can be obtained for $L = K^{-1}K$. $\square$

Proposition 4.5 *Let K be a Karle-Hauptman matrix, $K_{jk} = F_{H_j - H_k}$. Let $L = K^{-1}F$. Then*

$$L_l = \int_{\mathbf{V}} \rho^{-1}(r) \bar{\rho}(r) \exp[2\pi i H_l^T r] dr, \qquad (111)$$

where

$$\bar{\rho}(r) = \sum_{j=1}^{m} F_{H_j} \exp(-2\pi i H_j^T r). \qquad (112)$$

Proof. We show that $L_l = \sum_{j=1}^{m} [K^{-1}]_{lj} F_{H_j}$. By the definition of $\bar{\rho}(r)$,

$$L_l = \int_{\mathbf{V}} \rho^{-1}(r) \bar{\rho}(r) \exp[2\pi i H_l^T r] dr \qquad (113)$$

$$= \int_{\mathbf{V}} \rho^{-1}(r) \sum_{j=1}^{m} F_{H_j} \exp(-2\pi i H_j^T r) \exp[2\pi i H_l^T r] dr \qquad (114)$$

$$= \sum_{j=1}^{m} E_{H_l - H_j} F_{H_j} \qquad (115)$$

$$= \sum_{j=1}^{m} [K^{-1}]_{lj} F_{H_j}. \qquad (116)$$

$\square$

Proposition 4.6 *Let K be a Karle-Hauptman matrix, $K_{jk} = F_{H_j - H_k}$. Let $L = K^{-1}(F^* - F)$. Then*

$$L_l = \int_{\mathbf{V}} \rho^{-1}(r) \bar{\rho}(r) \exp[2\pi i H_l^T r] dr, \qquad (117)$$

where

$$\bar{\rho}(r) = \sum_{j=1}^{m} (F_{H_j}^* - F_{H_j}) \exp(-2\pi i H_j^T r) dr. \qquad (118)$$

18

Proof. Similar to the previous proposition, we show that

$$L_l = \sum_{j=1}^m [K^{-1}]_{lj}(F_{H_j}^* - F_{H_j}).$$

By the definition of $\bar{\rho}(r)$,

$$
\begin{aligned}
L_l &= \int_{\mathbf{V}} \rho^{-1}(r)\bar{\rho}(r)\exp[2\pi i H_l^T r]dr & (119)\\
&= \int_{\mathbf{V}} \rho^{-1}(r)\sum_{j=1}^m (F_{H_j}^* - F_{H_j})\exp(-2\pi i H_j^T r)\exp[2\pi i H_l^T r]dr & (120)\\
&= \sum_{j=1}^m E_{H_l-H_j}(F_{H_j}^* - F_{H_j}) & (121)\\
&= \sum_{j=1}^m [K^{-1}]_{lj}(F_{H_j}^* - F_{H_j}). & (122)
\end{aligned}
$$

$\square$

5 Concluding Remarks

In this paper, we studied the maximum entropy problem in the Bayesian statistical approach to the phase problem in protein X-ray crystallography. Since the solution of the problem is required in every step of the Bayesian method, an efficient algorithm for solving the problem is important especially for large-scale applications. Previous approaches used standard Newton's or approximation methods. They were either costly, requiring $\mathcal{O}(m^3)$ computation time, or not able to guarantee the fast convergence, where m is the number of structure factors of interest. We derived a formula to compute the inverse of the Hessian in $\mathcal{O}(m \log m)$ computation time, thereby reducing the time complexity of the Newton's method. With this formula, we will be able to apply the Newton's method to large-scale problems with low computation cost and fast covergence rate.

We described the entropy maximization problem and reviewed previous approaches to the problem. Some of the previous results were given only informally in literature. We gave more formal descriptions and provided accurate proofs for key mathematical facts. We think that this is necessary

for understanding the problem correctly and finding a riguous solution to it. In particular, we verified the regularity condition for the maximum entropy problem, which was neglected in the previous approaches. We studied the strong duality condition for the primal and dual entropy problems, and showed a necessary and sufficient condition for the dual minimum to be equal to the primal maximum. We also provided a proof for the positive definiteness of the Hessian matrix for the dual problem.

Our method for computing the inverse of the Hessian is based on the observation that the Hessian contains a positive definite matrix K and a rank-one update FF^H. Therefore, by using the Sherman-Morrison-Woodbury Theorem, the inverse of the Hessian can be computed as the inverse of the positive definite matrix K plus a simple matrix update. In the paper, we first developed a Sherman-Morrison-Woodbury formula for positive definite matrices, and then applied it to the Hessian matrix of the entropy problem to obtain an inverse update. We also showed that the inverse of the positive definite matrix K can be computed in $\mathcal{O}(m \log m)$ through a Fourier transform for the inverse of the electron density distribution.

Entropy maximization has broad applications. We only focused on its application in phase determination. Interested readers are referred to [14] for a general review on the subject. Here we would like to acknowledge Yunkai Zhou for bringing the reference [14] to our attention.

References

[1] Y. ALHASSID, N. AGMON, R. D. LEVIN, *An Upper Bound for the Entropy and Its Applications to the Maximal Entropy Problem*, Chem. Phys. Lett. Vol. 53, 1978, pp. 22–26.

[2] G. BRICOGNE, *Maximum Entropy and the Foundations of Direct Methods*, Acta Cryst. A40, 1984, pp. 410–445.

[3] G. BRICOGNE, *A Bayesian Statistical Theory of the Phase Problem. I. A Multichannel Maximum-Entropy Formalism for Constructing Generalized Joint Probability Distributions of Structure Factors*, Acta Cryst. A44, 1988, pp. 517–545.

[4] G. BRICOGNE, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. III. Extension to Powder Diffraction Data*, Acta Cryst. A47, 1991, pp. 803–829.

[5] G. BRICOGNE, *Direct Phase Determination by Entropy Maximization and Likelihood Ranking: Status Report and Perspectives*, Acta Cryst. D49, 1993, pp. 37–60.

[6] G. BRICOGNE, *Bayesian Statistical Viewpoint on Structure Determination: Basic Concepts and Examples*, in Methods in Enzymology, Vol. 276, 1997, Academic Press, pp. 361 – 423.

[7] G. BRICOGNE AND C. J. GILMORE, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. I. Theory, Algorithms and Strategy*, Acta Cryst. A46, 1990, pp. 284–297.

[8] H. E. DANIELS, *Saddlepoint Approximations in Statistics*, Ann. Math. Stat. 25, 1954, pp. 631–650.

[9] J. E. DENNIS, JR. AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1983.

[10] W. DONG, T. BAIRD, J. R. FRYER, C. J. GILMORE, D. D. MAC-NICOL, G. BRICOGNE, D. J. SMITH, M. A. O'KEEFE, AND S. HÖVMOLLER, *Electron Microscopy at 1-ÅResolution by Entropy Maximization and Likelihood Ranking*, Nature Vol. 355, 1992, pp. 605–609.

[11] S. DOUBLIE, S. XIANG, C. J. GILMORE, G. BRICOGNE, AND C. W. CARTER JR., *Overcoming Non-Isomorphism by Phase Permutation and Likelihood Scoring: Solution of the TrpRS Crystal Structure*, Acta Cryst. A50, 1994, pp. 164–182.

[12] S. DOUBLIE, G. BRICOGNE, C. J. GILMORE, AND C. W. CARTER JR., *Tryptophanyl-tRNA Synthetase Crystal Structure Reveals an Unexpected Homology to Tyrosyl-tRNA Synthetase*, Structure 15, 1995, 3:17–31.

[13] J. DRENTH, *Principles of Protein X-ray Crystallography*, Springer-Verlag, New York, New York, 1994.

[14] S. C. FANG, J. R. RAJASEKERA, AND H. S. J. TSAO, *Entropy Optimization and Mathematical Programming*, Kluwer Academic Publishers, Boston, Massachusetts, 1997.

[15] R. FLETCHER, *Practical Methods of Optimization*, John Wiley & Sons, New York, New York, 1987.

[16] E.L. FORTELLE AND G. BRICOGNE, *Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacememt and Multiwavelength Anomalous Diffraction Methods*, in Methods in Enzymology, Vol. 276, 1997, pp. 472 – 494.

[17] C. J. GILMORE, G. BRICOGNE, AND C. BANNISTER, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. II. Applications to Small Molecules*, Acta Cryst. A46, 1990, pp. 297–308.

[18] C. J. GILMORE, K. HENDERSON, AND G. BRICOGNE, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. IV. The Ab Initio Solution of Crystal Structures from Their X-ray Powder Data*, Acta Cryst. A47, 1991, pp. 830–841.

[19] C. J. GILMORE, K. HENDERSON, AND G. BRICOGNE, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. V. The Use of Likelihood as a Discriminator of Phase Sets Produced by the SAYTAN Program for a Small Protein*, Acta Cryst. A47, 1991, pp. 842–846.

[20] J. P. GLUSKER AND K. N. TRUEBLOOD, *Crystal Structure Analysis*, Oxford University Press, New York, New York, 1985.

[21] E. T. JAYNES, *Information Theory and Statistical Mechanics*, Phys. Rev. Vol. 106, pp. 620–630.

[22] E. T. JAYNES, *Prior Probabilities*, IEEE Trans. SSC-4, pp. 227–241.

[23] J. KARLE AND H. HAUPTMAN, *The Phases and Magnitudes of the Structure Factors*, Acta Cryst. 3, 1952, pp. 181–187.

[24] GEORGE PHILLIPS, RICHARD TAPIA, ZHIJUN WU, AND YIN ZHANG, *The Bayesian Statistical Approach to the Phase Problem in Protein X-ray Crystallography*, Technical Report, TR99-13, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1999.

[25] D. E. SANDS, *Introduction to Crystallography*, Dover Publications, Inc., New York, New York, 1975.

[26] K. SHANKLAND, C. J. GILMORE, G. BRICOGNE, AND HASHIZUME, *A Multisolution Method of Phase Determination by Combined Maximization of Entropy and Likelihood. VI. Automatic Likelihood Analysis via the Student t Test, with an Application to the Powder Structure of Magnesium Boron Nitride, Mg_3BN_3*, Acta Cryst. A49, 1993, pp. 493–501.

[27] A. H. SHERMAN, *On Newton Iterative Methods for the Solution of Systems of Nonlinear Equations*, SIAM J. Num. Anal., Vol. 15, 1978, pp. 755–771.

[28] ZHIJUN WU, GEORGE PHILLIPS, RICHARD TAPIA, AND YIN ZHANG, *The Bayesian Statistical Approach to the Phase Problem in Protein X-ray Crystallography*, Technical Report, TR99-13, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1999.

[29] S. XIANG, C. W. CARTER JR., G. BRICOGNE, AND C. J. GILMORE, *Entropy Maximization Constrained by Solvent Flatness: a New Method for Macromolecular Phase Extension and Map Improvement*, Acta Cryst. D49, 1993, pp. 193–212.