Molecular Conformations from Distance Matrices

W. Glunt, T.L. Hayden,* and M. Raydan

Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506

Received 31 March 1992; accepted 30 July 1992

Two algorithms are introduced that show exceptional promise in finding molecular conformations using distance geometry on nuclear magnetic resonance data. The first algorithm is a gradient version of the majorization algorithm from multidimensional scaling. The main contribution is a large decrease in CPU time. The second algorithm is an iterative algorithm between possible conformations obtained from the first algorithm and permissible data points near the configuration. These ideas are similar to alternating least squares or alternating projections on convex sets. The iterations significantly improve the conformation from the first algorithm when applied to the small peptide *E. coli* STh enterotoxin. © 1993 by John Wiley & Sons, Inc.

INTRODUCTION

Two methods that employ large-scale nonlinear optimization algorithms to predict the three-dimensional structures of biological molecules are potential energy methods and distance geometry. Recently, a truncated Newton method[1,2] was found to have several advantages over traditional methods for potential energy problems. In this article, we show that a recent gradient method introduced by Barzilai and Borwein[3] can reduce the CPU time for a majorization algorithm used in distance geometry by more than an order of magnitude. As the size of the molecules become large, the low storage requirement of a gradient method becomes the dominating issue.

Distance geometry methods[4–6] use constraints on the values of the interatomic distances to determine an ensemble of different conformations that satisfy the constraints. These methods are useful in the determination of molecular structure from nuclear magnetic resonance (NMR) data.

One important algorithm that has recently been adapted from multidimensional scaling to the conformation problem[7] is the majorization algorithm, which iteratively applys the Guttman transformation. The majorization algorithm appears to give superior results to previous embedding algorithms[5,8] but suffers from slow convergence. We will show that a gradient version of the algorithm converges much faster to the same function value in the examples from NMR data when the previous embedding algorithms are used to obtain good starting values.

The new gradient algorithm has the advantage that line searches are not necessary. Because the step length is related to the eigenvalues of the Hessian at the minimum, we will refer to the method as the spectral gradient method. We conjecture that this algorithm may also be competitive with other large-scale methods of nonlinear optimization currently used in chemical applications.

Because in distance geometry methods one is interested in finding those conformations that satisfy the distance constraints, we further introduce an iterative method to accomplish this goal. It is related to the alternating least-square methods used in multidimensional scaling.[9] Similar ideas of alternating projections[8,10] have been used to develop algorithms for matrix minimization problems associated with molecular conformation. These ideas of alternating projections on convex sets go back to Cheney and Goldstein.[11]

DISTANCE GEOMETRY METHODS

It is useful to think of the data being constrained in a rectangular parallelepiped we will call the data box B. The edges of the box are defined by upper, u_{ij}, and lower, l_{ij}, bounds so that the distance, d_{ij}, between atom i and atom j in a molecule with n atoms satisfies

$$l_{ij} \le d_{ij} \le u_{ij}, \qquad 1 \le i, j \le n,$$

where the matrix $[d_{ij}]$ of nonnegative distances is symmetrical with zero diagonal. The difference between the upper bound and a lower bound reflects the accuracy with which the data is known. To reflect this accuracy in the algorithms, it is important that weighted models be considered.

The goal of distance geometry is to find the ensemble of all conformations whose distances satisfy the bounds. A comprehensive approach to these

*Author to whom all correspondence should be addressed.

CCC 0192-8651/93/010114-07

problems is outlined by Havel in his DG-II[7] package. There are many procedures that use distance geometry in this program and we will only address one part, the so-called embedding stage. The first phase of the package uses bound smoothing[4,12] to tighten the bounds and we will assume this step has been achieved.

The Embedding Problem

The embedding problem asks for a conformation in a low-dimensional space such that the distances in the conformation approximate data that lie in the data box B. Results on the embedding problem using squared distances and an iterative projection algorithm[10] give improvements over earlier procedures,[5] which in essence use only the first step of the projection algorithm. The approximation algorithm should reflect the accuracy of the data by using large weights in those positions of the matrix where the difference in the upper and lower bounds are small. Extending the projection algorithms to weighted approximations has proved difficult. The use of a majorization algorithm that uses weighted approximation yields improved conformations over the unweighted approximations.[7] The majorization algorithm arises from multidimensional scaling, which has been considered in psychology and sociology for over 50 years. We will minimize a function closely related to the stress function in multidimensional scaling.

Stress

The notation and terminology of multidimensional scaling is followed to make the connection to the previous literature easily accessible. Column vectors are denoted by boldfaced lowercase Roman letters and matrices by capital Roman letters. Given the data box B, choose a set of numbers δ_{ij}, called dissimilarities, which obey the bounds.

$$0 \leq l_{ij} \leq \delta_{ij} \leq u_{ij}, \qquad i,j = 1, 2, \ldots, n$$

The resulting matrix $\Delta = [\delta_{ij}]$ is nonnegative symmetric with zero diagonal. We now seek n atoms (points) in three dimensions, denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, to form the molecular conformation with coordinates in an $n \times 3$ matrix X whose distances $d_{ij}(X)$ approximate δ_{ij}. The symmetric distance matrix $D(X)$, with elements $d_{ij}(X)$, contains the Euclidean distances between points in the conformation. Hence

$$d_{ij}^2(X) = (\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j).$$

$W = [w_{ij}]$ is a symmetric nonnegative matrix of weights with zero diagonal. The loss function, closely related to stress, to be minimized is

$$\sigma(X) = \frac{1}{2}\sum_i \sum_j w_{ij}(d_{ij}(X) - \delta_{ij})^2. \qquad (1)$$

When the normalization

$$\frac{1}{2}\sum_i \sum_j w_{ij}\delta_{ij}^2 = 1$$

is adopted, we have the form used by de Leeuw[13] to develop his majorization algorithm.

The Guttman Transform

If the function $\sigma(X)$ is differentiable at X,[13] then

$$\nabla\sigma(X) = 2(VX - B(X)X).$$

The elements of V are given by

$$v_{ij} = \begin{cases} -w_{ij} & \text{if } i \neq j \\ \Sigma_{k\neq i}\, w_{ik} & \text{otherwise.} \end{cases}$$

$B(X) = [b_{ij}(X)]$ is a function of X and is defined by

$$b_{ij}(X) = \begin{cases} -w_{ij}\delta_{ij}/d_{ij}(X) & \text{if } i \neq j \text{ and } d_{ij}(X) > 0 \\ 0 & \text{if } i \neq j \text{ and } d_{ij}(X) = 0 \\ -\Sigma_{k\neq i}\, b_{ik}(X) & \text{otherwise.} \end{cases}$$

The function $\sigma(X)$ is differentiable if and only if $d_{ij}(X) > 0$ for all i, j for which $w_{ij}\delta_{ij} > 0$. Because this holds in the chemistry applications, we have experienced no difficulty assuming differentiability. The more difficult subgradient case is discussed by de Leeuw.[14]

If one sets $\nabla\sigma(X) = 0$ and formally solves for X, we have

$$X = V^+B(X)X,$$

where V^+ is the Moore–Penrose inverse of V. The Guttman transformation is defined by

$$\Gamma(X) \equiv V^+B(X)X.$$

The Majorization Algorithm

In several articles,[13–15] de Leeuw and coauthors have shown that the following majorization algorithm yields good solutions of (1):

- Let $X_1 \in R^{n\times 3}$ be a starting configuration.
- While $\sigma(X_k) - \sigma(X_{k+1}) > \epsilon$
- $X_{k+1} = \Gamma(X_k)$
- End while.

The following list contains the major results for the algorithm:

- $\sigma(X_k)$ is nonincreasing and hence has a limit σ_∞.
- The proof of convergence does not assume differentiability and is established by clever use of Cauchy–Schwartz inequality.
- Each X_k is centered, that is, the centroid of the conformation is at the origin.

- The sequence X_k is bounded and hence has accumulation points denoted by S_∞.
- If $X_\infty \in S_\infty$, then $\sigma(X_\infty) = \sigma_\infty$.
- If $X_\infty \in S_\infty$ (assuming differentiability at X_∞), then X_∞ is stationary, that is, $X_\infty = \Gamma(X_\infty)$.
- If S_∞ is not a singleton, then it is a continuum.
- The algorithm is global, meaning that $\sigma(X_k)$ converges for each starting point but the value of σ_∞ and also S_∞ depends upon the starting value.
- If X_∞ is stationary, then for all 3×3 rotation matrices R $X_\infty R$ is also a stationary point and $\sigma(X_\infty R) = \sigma(X_\infty)$.
- All of the eigenvalues of the derivative of $\Gamma(X)$, denoted by Γ_X, are less than or equal to one.
- The rate of convergence of the algorithm is linear but may be slow if the largest eigenvalue (when the unit eigenvalues due to rotation are removed) of Γ_X is near one.

The following algorithm avoids the calculation of the Moore–Penrose inverse V^+. Further, it converges much faster than the majorization algorithm. Because we wish to change weights often in our applications, the avoidance of calculating V^+ is desirable. For the next algorithm, we require an inner product, $\langle A, B \rangle$, between two $n \times 3$ matrices A and B. Let

$$\langle A, B \rangle = \sum_{i=1}^{n} \sum_{j=1}^{3} a_{ij} b_{ij}.$$

The Spectral Gradient Method

- Let $X_0 \in R^{n \times 3}$ be a centered starting conformation.
- $X_1 = X_0 - \nabla_0(X_0) = X_0 - 2(VX_0 - B(X_0)X_0)$
- While $|\sigma(X_{k+1}) - \sigma(X_k)|/\sigma(X_k) > \epsilon$
- $S_{k-1} = X_k - X_{k-1}$
- $Y_{k-1} = \nabla\sigma(X_k) - \nabla\sigma(X_{k-1})$
- $\alpha_k = \langle S_{k-1}, Y_{k-1} \rangle / \langle S_{k-1}, S_{k-1} \rangle$
- $X_{k+1} = X_k - \alpha_k^{-1} \nabla\sigma(X_k)$
- End while.

ADVANTAGES:

- No line search is required (avoiding a time-intensive step).
- As with other gradient methods, the storage requirements are minimal.
- This algorithm is approximately 10–20 times faster than majorization.
- The Moore–Penrose inverse V^+ is not used.
- Each X_k is centered.

DISADVANTAGES:

- Convergence has only been established for strictly convex quadratic functions.
- The objective function does not monotonically decrease in general.
- The rate of convergence is linear and depends upon the spectrum of the Hessian at the local minimum.

STARTING CONFORMATIONS: Because the spectral algorithm is only local, a good starting value is necessary. However, the global majorization method also requires a good starting value to find a "good" local minimum. Using the following method for choosing a starting point, both methods have converged to the same value of σ_∞ in all of our molecular examples to date.

We begin by finding the matrix of squared distances D^2 that minimizes the Frobenius norm to the matrix $\Delta^2 = [\delta_{ij}^2]$. In general, the solution to this problem yields a conformation in a high-dimensional space. To find this conformation, carry out three or four iterations of the alternating projection algorithm, MAP, which is discussed in detail in references 8 and 10. The second step is to use the high-dimensional approximation to obtain a three-dimensional starting point. This is one of the first problems solved in this area[5,16] and the idea is to use the eigenvectors associated with the three largest eigenvalues of the high-dimensional result. The details are standard but may be found in the phase II portion of the same article explaining MAP.[8] One other step is desirable to obtain the starting point. We center the starting point by simply computing the centroid and subtracting this from the three-dimensional conformation. The forms of V and $B(X)$ then assure that the iterations in the spectral gradient algorithm are all centered. This is desirable in that translation invariance is removed and hence one is restricted to a subspace on which the Hessian of the objective function is positive definite at a local minimum while the Hessian is only positive semidefinite on the whole space. A detailed discussion of this issue may be found in the phase II development.[8]

We have not observed divergence of the algorithm in molecular applications using the above starting point. In this unlikely event, we suggest a few iterations of the Guttman transform from the original starting value to produce a better starting value for the spectral algorithm.

Relation to Previous Results

The earliest method to solve the unweighted eq. (1) was proposed by Kruskal[17] using the steepest descent method. Guttman[18] and de Leeuw and Heiser[13,15] observed that iterating the Guttman transform is equivalent to a gradient method with constant stepsize. It has also been observed that the Guttman transform iterations are related to an iterative power method. We will show in the next section that the spectral gradient method is related to a shifted power method. The step size is related to this shift and is one way of accounting for the increased rate of convergence.

LOCAL BEHAVIOR OF THE SPECTRAL GRADIENT METHOD

To study the local behavior of the spectral gradient method, we consider the minimization of a quadratic function $f(\mathbf{x}) = 1/2\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} + c$ with a symmetric and positive definite (SPD) Hessian matrix A. In this case, α_k becomes

$$\alpha_k = \frac{\mathbf{s}_{k-1}^T A\mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T\mathbf{s}_{k-1}}, \tag{2}$$

which is the Rayleigh quotient of A at the vector $\mathbf{s}_{k-1}$. Hence, it satisfies

$$0 < \lambda_{\min} \leq \alpha_k \leq \lambda_{\max} \qquad \text{for all } k \tag{3}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the smallest and largest eigenvalues of A. Lemma 1 demonstrates a connection between the spectral gradient method and the shifted power method to approximate eigenvectors and eigenvalues. This relationship will be used to explain the behavior of the method.

Lemma 1. *Let $f(\mathbf{x}) = 1/2\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} + c$ where A is a SPD matrix. Further, let $\mathbf{x}_\star$ be the unique minimizer of f, $\{\mathbf{x}_k\}$ the sequence generated by the spectral gradient method from a given vector $\mathbf{x}_0$, and $\mathbf{e}_k = \mathbf{x}_k - \mathbf{x}_\star$ for all k. Then*

1. $A\mathbf{e}_k = -\alpha_k\mathbf{s}_k$
2. $\mathbf{s}_{k+1} = \dfrac{1}{\alpha_{k+1}}(\alpha_k I - A)\mathbf{s}_k$

Proof: Because $\mathbf{x}_{k+1} = \mathbf{x}_k - 1/\alpha_k \nabla f(\mathbf{x}_k)$ and $\nabla f(\mathbf{x}_k) = A\mathbf{x}_k - \mathbf{b}$, where $\mathbf{b} = A\mathbf{x}_\star$, we have

$$A\mathbf{e}_k = -\alpha_k\mathbf{s}_k \qquad \text{for all } k. \tag{4}$$

Substituting $\mathbf{s}_k = \mathbf{e}_{k+1} - \mathbf{e}_k$ in (4), we obtain for any k

$$\mathbf{e}_{k+1} = \frac{1}{\alpha_k}(\alpha_k I - A)\mathbf{e}_k. \tag{5}$$

Combining (4) and (5), it follows that

$$\mathbf{s}_{k+1} = \frac{1}{\alpha_{k+1}}(\alpha_k I - A)\mathbf{s}_k \qquad \blacksquare$$

Because the scalars α_k satisfy (3) then, from (4), we conclude that the sequence $\{\mathbf{e}_k\}$ converges to zero if and only if the sequence $\{\mathbf{s}_k\}$ converges to zero. Therefore, for the minimization of a quadratic function with a SPD Hessian it suffices to study the behavior of the sequence of steps $\{\mathbf{s}_k\}$. For any given $\mathbf{s}_0$, there exist constants $c_1, c_2, \ldots, c_n$ such that

$$\mathbf{s}_0 = \sum_{i=1}^{n} c_i\mathbf{v}_i$$

where $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are orthonormal eigenvectors of A associated with the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. From Lemma 1, we can see that the generation of the sequence $\{\mathbf{s}_k\}$ resembles a shifted power method iteration, where the shift α_k is given by the Rayleigh quotient of A at the vector $\mathbf{s}_{k-1}$. In fact, for any integer k

$$\mathbf{s}_{k+1} = \frac{1}{\gamma_k}\sum_{i=1}^{n}\left(\prod_{j=0}^{k}(\alpha_j - \lambda_i)\right)c_i\mathbf{v}_i \tag{6}$$

where

$$\gamma_k = \prod_{j=1}^{k+1}\alpha_j.$$

From (6), we can see that if we use the exact eigenvalues of A as the scalars α_k in the spectral gradient method, in any order, then we find the exact solution in p iterations, where p is the number of distinct eigenvalues of A. At each iteration, we eliminate at least one coefficient from the eigenvector expansion (6). Unfortunately, the eigenvalues of the Hessian matrix A are not available. However, we can use the Rayleigh quotient α_k given by (2) to approximate the eigenvalues associated with the largest coefficients in the eigenvector expansion of $\mathbf{s}_{k-1}$ and obtain a significant reduction in the norm of $\mathbf{s}_{k+1}$. In fact, let us suppose that for some k

$$\mathbf{s}_{k-1} = \hat{c}_1\mathbf{v}_1 + \sum_{i=2}^{n}\epsilon_i\mathbf{v}_i$$

where

$$|\epsilon_i| \ll |\hat{c}_1|.$$

Then, by properties of the Rayleigh quotient, $\alpha_k = \lambda_1 + O(\epsilon^2)$ is an excellent estimate of λ_1. Therefore, the coefficient of $\mathbf{v}_1$ in the eigenvector expansion (6) of $\mathbf{s}_{k+1}$ will be greatly reduced relative to those of $\mathbf{v}_i$, $i = 2, \ldots, n$. After that, the sequence $\{\alpha_k\}$ will move toward a different eigenvalue. This process will continue until convergence. Recently, Raydan[19] established the following global convergence result.

Theorem 1. *Let $f(\mathbf{x})$ be a strictly convex quadratic function. Let $\{\mathbf{x}_k\}$ be the sequence generated by the spectral gradient method and $\mathbf{x}_\star$ the unique minimizer of f. Then, either $\mathbf{x}_j = \mathbf{x}_\star$ for some finite j or the sequence $\{\mathbf{x}_k\}$ converges to $\mathbf{x}_\star$.*

In their article, Barzilai and Borwein[3] presented a convergence analysis of this method only in the two-dimensional quadratic case, i.e., when the matrix A has only two distinct eigenvalues. For that particular case, they established R-superlinear convergence, an interesting feature. However, Fletcher[20] argued that, in general, only R-linear convergence should be expected.

Our experience, based upon extensive numerical results, seems to indicate that the rate of convergence of this method depends upon the distribution of the eigenvalues in the spectrum of A. In fact, the

spectral gradient method has a much better performance when the eigenvalues of the Hessian matrix A are clustered. This conclusion is consistent with the interpretation presented in the previous paragraphs.

DATA BOX SEARCHING

The method we propose has close ties to alternating least-squares methods used in multidimensional scaling[9,21] and alternating projections on convex sets as proposed by Cheney and Goldstein.[11] The goal is to find three-dimensional configurations whose distance matrices are in or near the data box B.

The iterative procedure for the search is the following. Begin with an initial set of dissimilarities $[\delta_{ij}^0]$ in the data box. By the spectral gradient algorithm, find a conformation X^1 whose distances $[d_{ij}(X^1)]$ approximate $[\delta_{ij}^0]$ in the weighted norm. Project the distance matrix $[d_{ij}(X^1)]$ onto the data box B. This projection produces a new set of dissimilarities $[\delta_{ij}^1]$ in the data box. The iteration is repeated until all distances $[d_{ij}(X^k)]$ are near $[\delta_{ij}^k]$ in the sense that their weighted differences converge to a value as near zero as possible. The projection on the box is straightforward.

Box Projection Algorithm. Let X^k be a three-dimensional conformation and $[d_{ij}(X)]$ the distance matrix generated by X^k. Then

$$\delta_{ij}^k = \begin{cases} d_{ij}(X^k) & \text{if } l_{ij} \le d_{ij}(X^k) \le u_{ij} \\ l_{ij} & \text{if } d_{ij}(X^k) < l_{ij} \\ u_{ij} & \text{if } d_{ij}(X^k) > u_{ij} \end{cases}$$

We denote this projection onto the data box B by $P_B(X^k)$.

The set of distance matrices generated by three-dimensional configurations is a manifold in the boundary of the cone of distance matrices that is unfortunately not a convex set. The cone of distance matrices and the low-dimensional faces of this cone are discused by Hayden et al.[22] Hence, when we apply the spectral gradient algorithm for a fixed set of dissimilarities we attempt to find the nearest distance matrix generated by all three-dimensional configurations. Because the minimizing set is not convex, this is not a standard projection method. In fact, there is no guarantee that the configuration found by the spectral gradient algorithm (or the majorization algorithm) is indeed the nearest one. However, the analogy to a projection method is strong and explains our choice of the notation $X = P_S([\delta_{ij}])$. We are now ready to present the search algorithm.

Data Box Algorithm

- Let $[\delta_{ij}^0]$ be an initial set of dissimilarities in B.
- While $\sigma^k > \epsilon$
- $X^k = P_S([\delta_{ij}^{k-1}])$
- $[\delta_{ij}^k] = P_B(X^k)$
- $\sigma^k = 1/2 \sum_i \sum_j w_{ij}(d_{ij}[X^k] - \delta_{ij}^k)^2$
- End while.

Starting Values. In order to compute X^k by the spectral gradient algorithm, one must use a starting value at each step k. For $k = 1$, use the starting value previously suggested. For $k > 1, X^{k-1}$ is a good starting value and saves $O(n^3)$ work.

Using the properties of P_S and P_B, one can show that the sequence σ^k is monotone decreasing. If the limit of σ^k is zero, then the limit configuration satisfies all the bounds.

NUMERICAL RESULTS

All the numerical experiments were run on an IBM 3090-600J using the VS FORTRAN compiler. The numerical methods are applied to find the conformation of the (active fragment) of the peptide *E. coli* STh enterotoxin, which inflicts dysentery. We use 14 residues of STh enterotoxin, which gives 160 atoms so that the distance matrix is 160 × 160. The data box is generated by the standard chemical bounds (generated in this case by DG-II[7]) and the NMR data from Ohkubo et al.[23] The initial bounds were smoothed by triangular and tetrangular methods in DG-II. The authors thank Tim Havel for making his program DG-II available. Using the smoothed data box bounds, the following procedure is applied to obtain each conformation:

1. Choose a random matrix $[\delta_{ij}]$ of dissimilarities in the data box B.
2. Obtain a starting configuration by applying four iterations of algorithm MAP.
3. Apply the spectral gradient algorithm to obtain a conformation $X^\star$.
4. Apply the data box algorithm to obtain a final configuration F.

We first report on the results of steps 1, 2, and 3. We use a fixed weight matrix, W, that is symmetrical with zero diagonal and

$$w_{ij} = \frac{1}{1 + 10(u_{ij} - l_{ij})}$$

The stopping value in the spectral gradient algorithm was set to 10^{-11}. To compare the CPU time (in seconds) between the majorization algorithm and the spectral gradient algorithm, we averaged the time required to find $X^\star$ in the first three steps for five different random starting dissimilarities. Both methods arrived at essentially the same $X^\star$ ($DME < 10^{-4}$; see below) from the same starting points. The average CPU time for majorization was 117.3 and 11.6 CPU for the spectral gradient method for a ratio of 10.1. We also ran MAP 50 times to find a starting

Table I. DME between five conformations from step 3.

Conformation	1	2	3	4	5
1	0	0.883	0.911	1.05	0.868
2		0	0.955	0.977	0.880
3			0	1.034	0.939
4				0	1.062
5					0

point (not as good; see ref. 8) and also obtained the same $X^\star$ for both methods. However, the majorization algorithm required an average of 800 CPU and the spectral gradient algorithm averaged 36 CPU for a ratio of 22. That the spectral gradient algorithm would both find the same solution as majorization and at an improved rate with an inferior starting value was unexpected.

One measure, the distance matrix error,[4] used to determine the dissimilarity between two conformations of the molecule is the following:

$$DME(d, \hat{d}) = \left\{ \frac{2}{n(n-1)} \sum_{i<j} (d_{ij} - \hat{d}_{ij})^2 \right\}^{1/2}.$$

For the five conformations (1, 2, ..., 5) found by step 3, the *DME* between pairs is reported in Table I.

We now report the results obtained by using all four steps. In the box algorithm, each step of a projection onto the box is interlaced with an application of step 3 to find a configuration $X^\star$. To reduce the total CPU time, the stopping criteria in step 3 was reduced from 10^{-11} to 10^{-6}. Because the method is related to alternating projections, high precision at each step is not crucial. The stopping criteria for step four was $\sigma < 10^{-5}$ and a maximal limit of 150 box projections was also enforced. In the data box, there were 533 positions so that $u_{ij} = l_{ij}$, which we call tight bounds, of a total of 25,440 bound restrictions. The same five random dissimilarities used in Table I were used as starting values to find five final conformations using all four steps. The following information on these five conformations is recorded in Table II. We record the time (in seconds) required to find each final configuration; the total number of bound violations (TBV) (the tight bounds are always slightly violated); the maximal bound violation (MBV) in Å; the average bound violation (AV); and the standard deviation of the bound violations (SD). We also record for the tight bounds the maximal tight violation (MTV), the average tight violation (ATV), and the standard tight deviation (STD).

Table III. DME between five conformations from step 4.

Conformation	1	2	3	4	5
1	0	1.24	1.39	1.49	1.1
2		0	1.38	1.33	1.21
3			0	1.39	1.38
4				0	1.36
5					0

As expected, the tight bounds have the largest weights and hence the violations for these positions is small. In Table III, we report the same information as in Table I except the final configurations are compared.

The DME is increased in Table III due to the following action of the algorithms. All the algorithms (MAP, spectral gradient, and box) tend to do a good job on the backbone of the peptide. This is partially due to the disulfide bonds that fix this part of the structure into a helical shape. MAP and one application of the spectral gradient algorithm tend to give structures in which some of the side chains are "compressed." This is due to the "projection" or embedding from a high-dimensional space into three dimensions. These compressed structures have relatively small DME with respect to each other, although the actual structure of each is not nearly as correct as the box structure, for example, ring structures are not well formed in one application of the spectral gradient algorithm but do appear in the box algorithm. This is a general trend in that the angles between atoms in the residues can be acute in the MAP or spectral gradient algorithm while these angles increase as the box algorithm is applied.

To compare some effects of different number of iterations of projections onto the data box, we record in Table IV information similar to Table II obtained from the same starting dissimilarity with the projections on the box carried out 30, 140, and 1000 times. Because of the excessive CPU time for 1000 iterations, this is only carried out on one conformation rather than five.

CONCLUSIONS

We have demonstrated that the weighted majorization algorithm may be replaced in some molecular

Table II. Time and bound violations for five conformations from step 4.

Conformation	Time	TBV	MBV	AV	SD	MTV	ATV	STD
1	311	688	0.51	0.027	0.038	0.12	0.02	0.02
2	269	706	0.52	0.029	0.038	0.12	0.02	0.02
3	465	699	0.27	0.027	0.037	0.15	0.02	0.02
4	267	687	0.61	0.028	0.04	0.12	0.02	0.02
5	300	677	0.55	0.028	0.04	0.17	0.02	0.02

Table IV. Time and bound violations for conformations generated by 30, 140, and 1000 iterations of box projections.

No. of iterations	Time	TBV	MBV	AV	SD	MTV	ATV	STD
30	65	766	0.81	0.073	0.09	0.5	0.05	0.02
140	341	688	0.5	0.027	0.038	0.12	0.02	0.02
1000	4033	654	0.13	0.011	0.015	0.08	0.008	0.009

conformation problems by the spectral gradient algorithm. The chief advantages are a magnitude of decrease in CPU time, low storage requirements, and that the gradient approach does not require calculation of the pseudoinverse of a singular matrix, which is required for the Guttman Transform. This last property allows one to change weights without excessive computation. The majorization algorithm has the advantage of global convergence, while the spectral gradient algorithm is only local (like Newton's method). However, in all molecular examples we examined both algorithms converge to the same conformation.

Preconditioning is known to greatly increase the speed of most optimization algorithms. For simple quadratic models, preconditioning significantly increases the rate of convergence of the spectral gradient algorithm[24] and makes it highly competitive with preconditioned conjugate gradient methods. We are presently investigating preconditioning for the spectral gradient algorithm for the problems considered in this article.

The second idea presented in this article is to alter the present search strategy for finding molecular conformations in the data box. Rather than fix a set of dissimilarities in the box and find the nearest conformation near (or in) the box to the given dissimilarities, we propose an alternate approach. Consider the three-dimensional configurations as a subset of the surface of the cone of all distance matrices. The data box may be regarded as a rectangular box. What is desired from a geometric viewpoint is to find the intersection of these two sets. Intuitively, our box algorithm begins with a set of dissimilarities in the box and finds a nearby configuration whose generated distances "best" approximate the dissimilarities. Because the generated distances are usually outside the box, these distances are projected onto the box and a new set of dissimilarities is obtained. The process is then repeated. The effect is that one is "projecting" back and forth between the two sets. On each projection, one hopes the distance between the points in the sets is decreased. Our numerical results indicate this is a fruitful idea, and indeed a proof can be given if both sets are convex.

We find that repeated projections do tend toward each other. However, to completely remove all bound violations requires excessive CPU time under the simple selection of weights proposed. On the other hand, the present box algorithm will find excellent starting values for other algorithms. We are presently experimenting with other choices of weights and also propose to change the weights between certain iterations of the box algorithm. With these modifications, the box algorithm should compete with more complicated algorithms now in use, such as simulated anneling.

W.G. was supported by NSF Grant CHE-9005960, T.L.H. was supported by NSF Grant CHE-9005960, and M.R. was partially supported by Central University of Venezuela.

References

1. T. Schlick and M. Overton, *J. Comp. Chem.*, **8,** 1025 (1987).
2. T. Schlick, C. Peskin, S. Broyde, and M. Overton, *J. Comp. Chem.*, **8,** 1199 (1987).
3. J. Barzilai and J. Borwein, *IMA J. Numer. Anal.*, **8,** 141 (1988).
4. G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, Wiley, New York, 1988.
5. T. Havel, I. Kuntz, and G. Crippen, *Bull. Math. Biol.*, **45,** 665 (1983).
6. T. Havel and K. Wüthrich, *Bull. Math. Biol.*, **46,** 699 (1984).
7. T.F. Havel, *Prog. Biophys. Mol. Biol.*, **56,** 43 (1991).
8. W. Glunt, T.L. Hayden, and W.-M. Liu, *Bull. Math. Biol.*, **53,** 769 (1991).
9. Y. Takane, F. Young, and J. de Leeuw, *Psychometrika*, **42,** 7 (1977).
10. W. Glunt, T.L. Hayden, S. Hong, and J. Wells, *SIAM J. Matrix Anal. Appl.*, **11,** 589 (1990).
11. W. Cheney and A. Goldstein, *Proc. Am. Math. Soc.*, **10,** 443 (1959).
12. P. Easthope and T. Havel, *Bull. Math. Biol.*, **51,** 173 (1989).
13. J. de Leeuw, *J. Classification*, **5,** 163 (1988).
14. J. de Leeuw, In *Recent Developments in Statistics*, J.R. Barra et al., Eds., North-Holland, Amsterdam, 1977, pp. 133–145.
15. J. de Leeuw and W. Heiser, In *Multivariate Analysis—V*, P.R. Krishnaiah, Ed., North-Holland, Amsterdam, 1980, pp. 501–522.
16. C. Eckart and G. Young, *Psychometrika*, **1,** 211 (1936).
17. J.B. Kruskal, *Psychometrika*, **29,** 115 (1964).
18. L. Guttman, *Psychometrika*, **33,** 469 (1968).
19. M. Raydan, *IMA J. Numer. Anal.*, to appear.
20. R. Fletcher, *Lectures Appl. Math.*, **26,** 165 (1990).
21. J. Meulman, *A Distance Approach to Nonlinear Multivariate Analysis*, DSWO Press, Leiden, The Netherlands, 1986.
22. T.L. Hayden, J. Wells, W. Liu, and P. Tarazaga, *Linear Algebra Appl.*, **144,** 153 (1991).
23. T. Ohkubo, Y. Kobayashi, Y. Shimonishi, Y. Kyogoku, W. Braun, and N. Go, *Biopolymers*, **25,** S123 (1986).
24. M. Raydan, PhD thesis, Rice University, Houston, TX, 1991.