

ARGONNE NATIONAL LABORATORY  
9700 South Cass Avenue  
Argonne, Illinois 60439

**GLOBAL SMOOTHING AND CONTINUATION FOR LARGE-SCALE  
MOLECULAR OPTIMIZATION**

**Jorge J. Moré and Zhijun Wu**

Mathematics and Computer Science Division

Preprint MCS-P539-1095

October 1995

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38, and by the Argonne Director's Individual Investigator Program.

# GLOBAL SMOOTHING AND CONTINUATION FOR LARGE-SCALE MOLECULAR OPTIMIZATION

Jorge J. Moré and Zhijun Wu

## Abstract

We discuss the formulation of optimization problems that arise in the study of distance geometry, ionic systems, and molecular clusters. We show that continuation techniques based on global smoothing are applicable to these molecular optimization problems, and we outline the issues that must be resolved in the solution of large-scale molecular optimization problems.

## 1 Introduction

We are concerned, in particular, with molecular optimization problems that arise in the study of protein structures in biophysical chemistry. If we adopt the hypothesis that the native protein structure corresponds to the global minimum of the protein energy [43, 48], then the protein structure is determined by minimizing a potential energy function in conformational space. If the protein structure is determined from bounds on distances between pairs of atoms and other geometric constraints on the protein, then distance geometry [9, 20] techniques are required. Both approaches require the solution of global optimization problems.

The problem of finding the least energy structure for a given molecular system arises not only in biological studies, but also in the study of large, confined ionic systems in plasma physics [19, 39, 44]. Configurations of systems with up to 200,000 ions are of special interest because phase transitions occur for systems of this size. Such configurations require determining the global minimum of a function with up to 600,000 variables — a computationally intensive problem, even for local minimization, because the problem is dense and each function evaluation requires order  $n^2$  floating-point operations (flops).

Molecular optimization problems also arise in the study of clusters; for an introduction to the problems in this area, see the books edited by Reynolds [41] and Haberland [18]. Much of the interest in clusters is due to unexpected theoretical and practical results, such as the discovery of the stable carbon cluster  $C_{60}$ . Theoretical properties of clusters usually are determined by molecular dynamics simulation or by potential energy minimization. Small argon clusters have received considerable attention in the past (see, for example, the classical studies of Hoare [23] and Northby [37]), since they have simple potentials and structure, but current interest centers on clusters with more involved potentials.

---

This work was supported by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Computational and Technology Research, U.S. Department of Energy, under Contract W-31-109-Eng-38, and by the Argonne Director's Individual Investigator Program.

The molecular optimization problems that we have mentioned are difficult because the presence of a large number of local minimizers, even for systems with a small number of atoms, creates numerous regions of attraction for local searches. We have been using global smoothing and continuation techniques for these problems. In this approach the Gaussian transform is used to map the objective function into a smoother function with fewer local minimizers, and an optimization procedure is applied to the transformed function, tracing the minimizers as the function is gradually changed back to the original function. A transformed function is a coarse approximation to the original function, with small and narrow minimizers being removed while the overall structure is maintained. This property allows the optimization procedure to skip less interesting local minimizers and to concentrate on regions with average low function values, where a global minimizer is most likely to be located.

Global smoothing was first proposed in the diffusion equation method for protein conformation by Scheraga and coworkers [38, 27, 28, 29, 43]. A similar approach was used in the packet annealing algorithm of Shalloway [48, 47] and in the algorithms of Coleman, Shalloway, and Wu [7, 8] for molecular conformation problems. Recent developments of global smoothing and continuation techniques from a mathematical and computational point of view can be found in Wu [51] and Moré and Wu [35, 34], while Moré and Wu [36] have extended the smoothing properties of the Gaussian transform to transformations with other density functions.

While global smoothing is a promising approach for the solution of molecular optimization problems, many theoretical and computational issues need additional attention. In this paper we review current work on distance geometry problems, and indicate how these results can be extended to ionic systems and molecular clusters.

In Section 2 we discuss the formulation of molecular optimization problems that arise in distance geometry calculations, ionic systems, and molecular clusters. Section 3 is a review of the properties of the Gaussian transform. In particular, we outline the techniques used to compute the Gaussian transform for molecular optimization problems. In Section 4 we discuss computational experiments carried out with a simple continuation algorithm, and we show how problem formulation affects the choice of optimization procedure in the continuation algorithm. We conclude the paper in Section 5 with a critical review of issues that must be addressed in order to solve large-scale molecular optimization problems on high-performance architectures. We address, in particular, the evaluation of the function by fast multipole methods.

## 2 Molecular Optimization Problems

A typical molecular optimization problem is to determine a structure with minimal potential energy. In some cases, the position of the atoms in the structure must also satisfy certain

physical constraints. In this section we review three molecular optimization problems and related work.

## 2.1 Distance Geometry

Distance geometry problems arise in the interpretation of nuclear magnetic resonance (NMR) data and in the determination of protein structures. For a general review of the distance geometry problem and its relationship to macromolecular modeling, see Crippen and Havel [9], Havel [20], Kuntz, Thomason, and Oshiro [30], and Brünger and Nilges [4].

A distance geometry problem is specified by a subset  $\mathcal{S}$  of all atom pairs and by the distances  $\delta_{i,j}$  between atoms  $i$  and  $j$  for  $(i, j) \in \mathcal{S}$ . A solution to the distance geometry problem is a set of positions  $x_1, \dots, x_m$  in  $\mathbb{R}^3$  such that

$$\|x_i - x_j\| = \delta_{i,j}, \quad (i, j) \in \mathcal{S}. \quad (2.1)$$

Usually,  $\mathcal{S}$  is sparse; in other words, only a small subset of distances is known.

In practice, lower and upper bounds on the distances are specified instead of their exact values. The distance geometry problem with lower and upper bounds is to find positions  $x_1, \dots, x_m$  such that

$$l_{i,j} \leq \|x_i - x_j\| \leq u_{i,j}, \quad (i, j) \in \mathcal{S}, \quad (2.2)$$

where  $l_{i,j}$  and  $u_{i,j}$  are lower and upper bounds on the distance constraints, respectively. An important case of this problem is to obtain an  $\varepsilon$ -optimal solution to the distance geometry problem, that is, positions  $x_1, \dots, x_m$  such that

$$\left| \|x_i - x_j\| - \delta_{i,j} \right| \leq \varepsilon, \quad (i, j) \in \mathcal{S} \quad (2.3)$$

for some  $\varepsilon > 0$ . An  $\varepsilon$ -optimal solution is useful when the exact solution to the problem (2.1) does not exist because of small errors in the data. This situation can happen, for example, when the triangle inequality

$$\delta_{i,j} \leq \delta_{i,k} + \delta_{k,j}$$

is violated for atoms  $\{i, j, k\}$  because of possible inconsistencies in the experimental data.

The distance geometry problem (2.1) is computationally intractable because the restriction of the distance geometry problem to a one-dimensional space is equivalent to the set partition problem, which is known to be NP-complete [10]. Saxe [42] shows that  $k$ -dimensional distance geometry problems are strongly NP-hard for all  $k \geq 1$ . The following result of Moré and Wu [34] shows that obtaining an approximate solution to the distance geometry problem is also NP-hard.

**Theorem 2.1** *Determining an  $\varepsilon$ -optimal solution to the distance geometry problem in  $\mathbb{R}$  is NP-hard.*

The distance geometry problems that we have described can be formulated as global optimization problems for which the constraints are satisfied at a global minimizer of the problem. A simple formulation is in terms of finding the global minimum of the function

$$f(x) = \sum_{i,j \in \mathcal{S}} p_{i,j}(x_i - x_j), \quad (2.4)$$

where the pairwise potential  $p_{i,j} : \mathbb{R}^n \mapsto \mathbb{R}$  is defined for problem (2.1) by

$$p_{i,j}(x) = \left( \|x\|^2 - \delta_{i,j}^2 \right)^2, \quad (2.5)$$

while Crippen and Havel [9] suggested that for problem (2.2)

$$p_{i,j}(x) = \min^2 \left\{ \frac{\|x\|^2 - l_{i,j}^2}{l_{i,j}^2}, 0 \right\} + \max^2 \left\{ \frac{\|x\|^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\}. \quad (2.6)$$

Clearly,  $x = \{x_1, \dots, x_m\}$  solves the distance geometry problem if and only if  $x$  is a global minimizer of  $f$  and  $f(x) = 0$ .

Special optimization algorithms have been developed for solving the distance geometry problem (2.1). For example, Hendrickson [21, 22] used a graph-theoretic viewpoint to develop algorithms that test the uniqueness and rigidity of the distance graph. These algorithms can be used to reduce the problem into smaller, easier subproblems. Glunt, Hayden, and Raydan [12, 13] have proposed a special gradient method for determining a local minimizer of the problem defined by (2.4) with

$$p_{i,j}(x) = (\|x\| - \delta_{i,j})^2.$$

Al-Homidan and Fletcher [1] have done related work on a hybrid algorithm that combines an alternating projection method with a quasi-Newton method.

If all pairwise distances are known and a solution exists, then the solution of the distance geometry problem (2.1) can be determined (Blumenthal [3, Section 43], Crippen and Havel [9, Section 6.3]) by computing the largest three eigenvalues and eigenvectors of the rank-3 positive semidefinite matrix  $A \in \mathbb{R}^{m \times m}$  defined by

$$a_{i,j} = \frac{1}{2} \left( \delta_i^2 + \delta_j^2 - \delta_{i,j}^2 \right), \quad i, j = 1, \dots, m,$$

where  $\delta_i = \|x_i - x_0\|$ , and  $x_0 \in \mathbb{R}^n$  is a convex combination of  $x_1, \dots, x_m$ . Note that if  $x_0$  is a convex combination of  $x_1, \dots, x_m$ , then  $\delta_i$  can be expressed in terms of  $\delta_{i,j}$ . In practice,  $x_0$  is the centroid of  $x_1, \dots, x_m$ . We can determine the coordinates  $x_1, \dots, x_m$  by noting that the identity

$$2(x_i - x_0)^T(x_j - x_0) = \|x_i - x_0\|^2 + \|x_j - x_0\|^2 - \|x_i - x_j\|^2$$

implies that  $A = B^T B$  is a positive semidefinite rank-3 matrix with

$$B = (x_1 - x_0, \dots, x_m - x_0).$$

Hence, the vectors  $x_1, \dots, x_m$  can be determined by computing the largest three eigenvalues and eigenvectors of the rank-3 matrix  $A$ . Alternatively, we could use the Cholesky decomposition with diagonal pivoting.

In practice, only a small subset of the distances is known, and there are experimental errors in the data, so the above procedure cannot be used. However, an extension of this procedure is employed by the **embed** algorithm (see Crippen and Havel [9], and Havel [20]) in practical distance geometry calculations. In the first phase of the **embed** algorithm, the sparse set of distance constraints is extended by using the relationships

$$u_{i,j} = \min(u_{i,j}, u_{i,k} + u_{k,j}), \quad l_{i,j} = \max(l_{i,j}, l_{i,k} - u_{k,j}, l_{j,k} - u_{k,i}).$$

Given a full set of bounds, distances  $\delta_{i,j} \in [l_{i,j}, u_{i,j}]$  are chosen, and an attempt is made to compute coordinates  $x_1, \dots, x_m$  as in the above procedure. This attempt usually fails, but it can be used to generate a rank-3 approximation to  $A$ , which leads to an approximation to the solution of problem (2.1). This approximation can be refined by minimizing a function of the form (2.4,2.5).

The **embed** algorithm, as described above, may require many trial choices of  $\delta_{i,j}$  in  $[l_{i,j}, u_{i,j}]$  before a solution to problem (2.2) is found. Current implementations of the **embed** algorithm use a local minimizer of the problem defined by (2.4) and (2.5) as a starting point for a simulated annealing procedure. In Section 4 we will outline the proposal of Moré and Wu [34] for finding a solution of the distance geometry problem (2.2) by directly determining a global minimizer of the function defined by (2.4) and (2.6).

## 2.2 Ionic Systems

The potential energy for a confined ionic system of  $m$  ions located at  $x_1, \dots, x_m$  in  $R^3$  can be modeled, for example, by a function of the form

$$f(x) = \sum_{i \neq j} p_{i,j}(x_i - x_j) + \sum_{i=1}^m \|x_i\|^2, \quad (2.7)$$

where  $p_{i,j} : \mathbb{R}^3 \mapsto \mathbb{R}$  is defined by

$$p_{i,j}(x) = v(\|x\|), \quad v(r) = r^{-1}. \quad (2.8)$$

Hasse and Schiffer [19], Rafac, Schiffer, Hangst, Dubin, and Wales [39], and Schiffer [44] studied configurations of confined ionic systems via molecular dynamics simulation. The results for small systems showed that ionic systems have a layered shell structures, with the

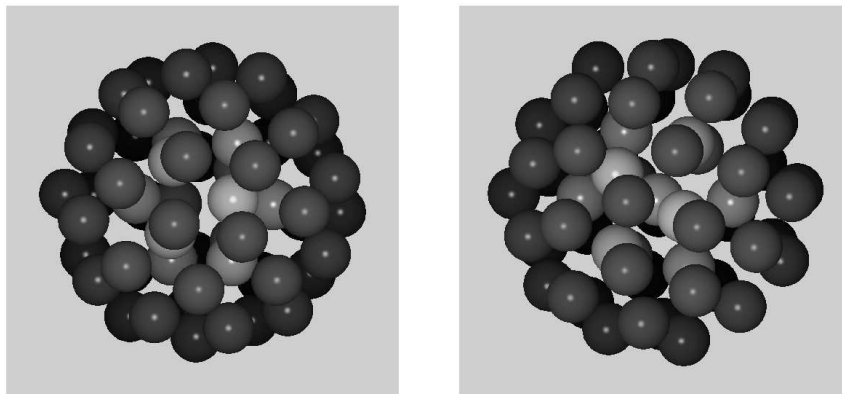


Figure 2.1: Confined ionic system with 60 ions (left) and 61 atoms (right).

number of shells increasing as the number of atoms in the system increases. The distribution of the ions over the shells also varies with different systems. For example, Figure 2.1 shows that the system of 60 ions has two shells with 12 ions in the inner shell and 48 ions in the outer shell, while the system of 61 ions has three shells with a single ion (at the center of the system) as the innermost shell.

Experiments indicate that as the number of atoms increases, the boundaries between the shells become blurred, and eventually the system achieves a crystal form. Therefore, there must be a phase transition from a system of layered shells to a body-centered cubic lattice, a standard crystal structure. This phase transition is of special physical interest, but in order to locate the transition, configurations for very large systems (say 200,000 ions) must be determined. This calculation would be prohibitively expensive for a molecular dynamics simulation, but may be possible by minimizing the potential energy function. We have found in our preliminary studies that configurations for most systems with up to 100 ions can be determined by a single local minimization. Of course, for large systems, global optimization algorithms are required to obtain the most stable configurations.

### 2.3 Molecular Clusters

A cluster is a group of identical molecules with specific geometrical and chemical properties. Clusters of chemical importance include, for example, argon and carbon clusters. A fundamental problem in cluster science is to determine the geometrical structure of clusters in their lowest energy states. Related problems include structure changes from clusters to

bulk matters and low energy paths between stable states. For a general review of these topics, see Haberland [18].

Clusters of argon molecules were first studied by Hoare and coworkers [25, 24, 23]. Northby [37] obtained the structures for clusters with up to 147 molecules using a lattice search algorithm, which later was improved and used for even larger clusters by Xue [52]. Results for small argon clusters have also been obtained by general-purpose algorithms such as the diffusion equation method [28], the packet annealing algorithm [48], the stochastic search method [6, 5], and the effective energy simulated annealing algorithm [7].

Argon clusters have been heavily studied because the potentials and structure of these clusters are relatively simple. Argon clusters usually are modeled by the Lennard-Jones potential

$$f(x) = \sum_{i \neq j} p_{i,j}(x_i - x_j),$$

where  $p_{i,j} : \mathbb{R}^3 \mapsto \mathbb{R}$  is defined by

$$p_{i,j}(x) = v(\|x\|), \quad v(r) = r^{-12} - 2r^{-6},$$

or by the Morse potential

$$v(r) = (1 - \exp[\alpha(1 - r)])^2 - 1,$$

for some positive constant  $\alpha$ . For example, Hoare [23] used  $\alpha = 3$ .

Potentials for other clusters can be more involved. For example, in the study of metal clusters [41, 18, 26] it is common to use potentials of the form

$$f(x) = \sum_{i \neq j} p_{i,j}(x_i - x_j) - \sum_{j=1}^n \left( \sum_{i \neq j} q_{i,j}(x_i - x_j) \right)^{1/2},$$

where  $p_{i,j} : \mathbb{R}^3 \mapsto \mathbb{R}$  and  $q_{i,j} : \mathbb{R}^3 \mapsto \mathbb{R}$  are of the form  $v(\|x\|)$  with

$$v(r) = \alpha \exp[-\beta(r - 1)],$$

for positive constants  $\alpha$  and  $\beta$ . Note that these potentials are functions of the pairwise distance between atoms, and that they decay rapidly as  $r$  approaches infinity. The potential for ionic systems, on the other hand, decays slowly as  $r$  approaches infinity.

Cluster problems are difficult for most global optimization strategies because they tend to have a large number of local minimizers that act as points of attraction for any local minimizer. For argon clusters, Hoare [23] found that systems with  $6 \leq m \leq 13$  atoms had

$$2, 4, 8, 18, 57, 145, 366, 989$$

different minima, respectively, and on the basis of this observation conjectured that the number of minima grew like  $\exp(m^2)$ . We are usually interested in global minimizers, but local minimizers with low function values are also of interest because they represent the most stable structures. For a discussion of these issues see, for example, Jellinek [26].



### 3 Smoothing Transformations

The global continuation approach to finding the global minimizer is to transform the function into a smoother function with fewer local minimizers, apply an optimization algorithm to the transformed function, and trace the minimizers back to the original function. This approach is well suited for problems with many local minimizers. As already noted, molecular optimization problems tend to have a large number of local minimizers.

A transformed function is a coarse approximation to the original function, with small and narrow minimizers being removed, while the overall structure of the function is maintained. This property allows the optimization algorithm to skip less interesting local minimizers and to concentrate on regions with average low function values, where a global minimizer is most likely to be located.

The smoothing transform, called the Gaussian transform, depends on a parameter  $\lambda$  that controls the degree of smoothing. The original function is obtained if  $\lambda = 0$ , while smoother functions are obtained as  $\lambda$  increases.

**Definition 3.1** *The Gaussian transform  $\langle f \rangle_\lambda$  of a function  $f : \mathbb{R}^n \mapsto \mathbb{R}$  is*

$$\langle f \rangle_\lambda(x) = \frac{1}{\pi^{n/2} \lambda^n} \int_{\mathbb{R}^n} f(y) \exp\left(-\frac{\|y - x\|^2}{\lambda^2}\right) dy. \quad (3.1)$$

The value  $\langle f \rangle_\lambda(x)$  is an average of  $f$  in a neighborhood of  $x$ , with the relative size of this neighborhood controlled by the parameter  $\lambda$ . The size of the neighborhood decreases as  $\lambda$  decreases so that when  $\lambda = 0$ , the neighborhood is the center  $x$ . The Gaussian transform  $\langle f \rangle_\lambda$  can also be viewed as the expected value of  $f$  with respect to the Gaussian density function

$$\rho_\lambda(y) = \frac{1}{\pi^{n/2} \lambda^n} \exp\left(-\frac{\|y\|^2}{\lambda^2}\right).$$

For the mathematical properties of the Gaussian transform, readers are referred to Wu [51] and Moré and Wu [35]. We also note that other density functions may be considered; an analysis of these generalized transformations can be found in Moré and Wu [36].

Motivation for the Gaussian transform can be obtained by showing that the Gaussian transform of the two-dimensional version of the Griewank function

$$f(x) = 1 + \sum_{i=1}^n \left( \frac{x_i^2}{200} \right) - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad (3.2)$$

removes local minimizers. This function was constructed by Griewank [16] to test global optimization algorithms on problems with a large number of local minimizers. Figure 3.1 shows plots of the Griewank function and the Gaussian transform

$$\langle f \rangle_\lambda(x) = 1 + \sum_{i=1}^n \left( \frac{x_i^2}{200} + \frac{\lambda^2}{400} \right) - \prod_{i=1}^n \exp\left(-\frac{\lambda^2}{4i}\right) \cos\left(\frac{x_i}{\sqrt{i}}\right) \quad (3.3)$$

of the Griewank function; justification for (3.3) as the Gaussian transform of the Griewank function will be provided shortly.

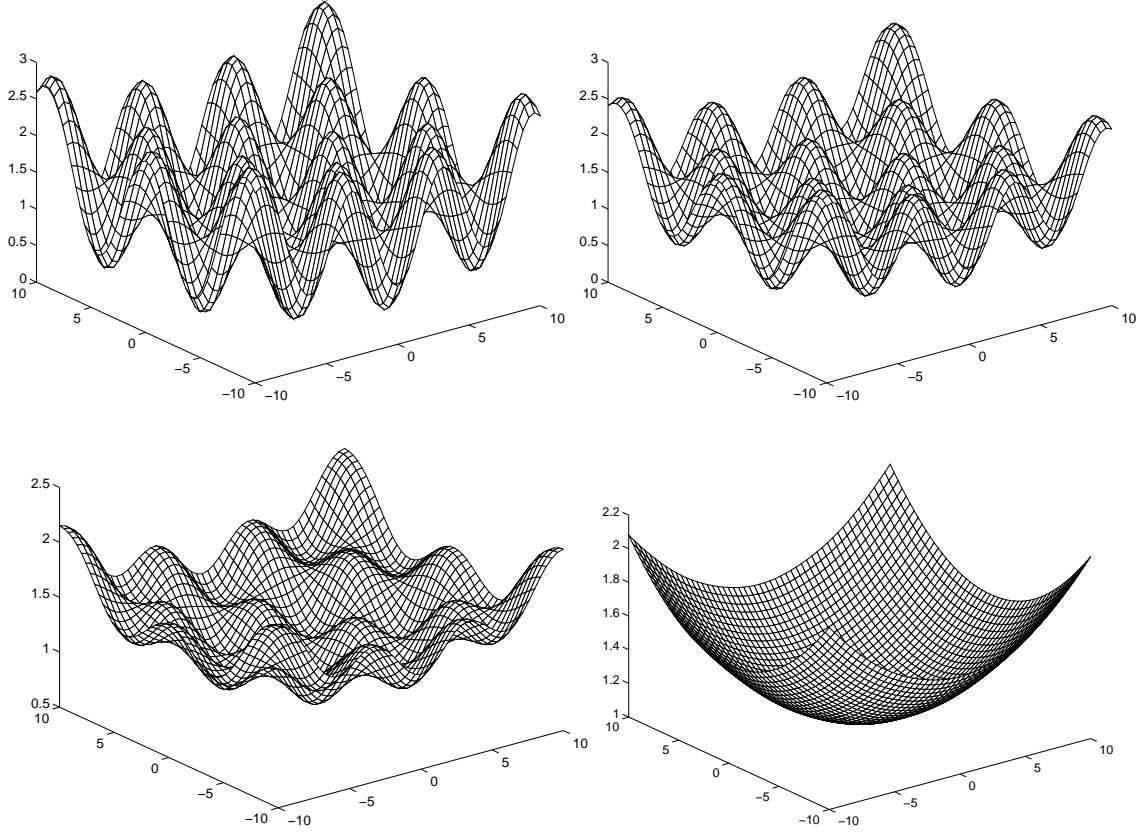


Figure 3.1: The Gaussian transform for the Griewank function. The original function ( $\lambda = 0$ ) is in the top left corner, with  $\lambda = 1$  in the top right corner,  $\lambda = 2$  in the bottom left corner, and  $\lambda = 4$  in the bottom right corner.

Figure 3.1 shows that the Gaussian transform reduces the number of minimizers as  $\lambda$  increases, and that the global minimizer of the original function can be found by applying a local minimization algorithm to the transformed functions and tracing the minimizers back to the original function. Although these plots are suggestive, it is important to keep in mind that the global smoothing approach is not guaranteed to succeed in all cases.

The following result of Wu [51] explains why the Gaussian transform reduces the high-frequency components of the function and eliminates local minimizers as  $\lambda$  increases.

**Theorem 3.2** *If  $\widehat{f} : \mathbb{R}^n \mapsto \mathbb{C}$  is the Fourier transform of  $f : \mathbb{R}^n \mapsto \mathbb{R}$ , then*

$$\left| \widehat{\langle f \rangle_\lambda}(\omega) \right| = \exp \left( -\frac{1}{4} \lambda^2 \|w\|^2 \right) \left| \widehat{f}(\omega) \right|.$$

Theorem 3.2 shows that  $\widehat{\langle f \rangle_\lambda}(\omega)$ , the component of  $\langle f \rangle_\lambda$  for frequency  $\omega$ , can be made

arbitrarily small if  $\lambda\|\omega\|$  is sufficiently large. In particular, the high-frequency components are removed when  $\lambda$  is large.

Computing the Gaussian transform usually requires the evaluation of  $n$ -dimensional integrals, but for many functions that arise in practice, it is possible to compute the Gaussian transform explicitly in terms of one-dimensional transforms. In particular, if the function is *decomposable*, that is, if the function  $f : \mathbb{R}^n \mapsto \mathbb{R}$  can be written in the form

$$f(x) = \sum_{k=1}^m \left( \prod_{j=1}^n f_{k,j}(x_j) \right),$$

for some set of functions  $\{f_{k,j}\}$ , where  $f_{k,j} : \mathbb{R} \mapsto \mathbb{R}$ , then a simple computation shows that

$$\langle f \rangle_\lambda(x) = \sum_{k=1}^m \left( \prod_{j=1}^n \langle f_{k,j} \rangle_\lambda(x_j) \right).$$

Thus, computing  $\langle f \rangle_\lambda$  for a decomposable function requires the computation of only one-dimensional transforms  $\langle f_{k,j} \rangle_\lambda$ .

The Gaussian transform of polynomials and other analytic functions have been determined by Kostrowicki and Piela [27] by using an alternative definition of the Gaussian transform. For a derivation that uses definition (3.1), see Moré and Wu [35]. These results show, in particular, that

$$t \mapsto t^2 + \frac{1}{2}\lambda^2, \quad t \mapsto \cos(\alpha t) \exp\left(-\frac{1}{4}\alpha^2\lambda^2\right),$$

are the Gaussian transforms of the functions  $t \mapsto t^2$  and  $t \mapsto \cos(\alpha t)$ , respectively. Since (3.2) is decomposable, the Gaussian transform of the Griewank function is given by (3.3).

For distance geometry problems, as well as for many other molecular problems, we are interested in transforming a class of functions defined in terms of the distances between pairs of atoms. Given functions  $p_{i,j} : \mathbb{R}^p \mapsto \mathbb{R}$  of the distances between atoms  $i$  and  $j$ , we consider a general function

$$f(x) = \sum_{i,j \in \mathcal{S}} p_{i,j}(x_i - x_j), \tag{3.4}$$

where  $\mathcal{S}$  is some subset of all pairs of atoms, and  $x_i \in \mathbb{R}^p$  is the position of the  $i$ -th atom. We are concerned with three-dimensional problems where  $p = 3$ .

The following result of Moré and Wu [35] shows that computing the Gaussian transform of (3.4) requires only the Gaussian transform of  $p_{i,j}$ .

**Theorem 3.3** *If  $f : \mathbb{R}^n \mapsto \mathbb{R}$  and  $h : \mathbb{R}^p \mapsto \mathbb{R}$  are related by*

$$f(x) = h(P^T x),$$

*for some matrix  $P \in \mathbb{R}^{n \times p}$  such that  $P^T P = \sigma^2 I$ , then*

$$\langle f \rangle_\lambda(x) = \langle h \rangle_{\sigma\lambda}(P^T x).$$

Theorem 3.3 reduces the computation of the Gaussian transform of  $f$ , which is defined on  $\mathbb{R}^n$ , to the computation of the Gaussian transform of  $h$ , which is defined on  $\mathbb{R}^p$ . As an application of this result, note that

$$\langle f \rangle_\lambda(x) = \sum_{i,j \in \mathcal{S}} \langle p_{i,j} \rangle_{\sqrt{2}\lambda}(x_i - x_j)$$

is the Gaussian transform for the function defined by (3.4). In this case  $f$  is defined on  $\mathbb{R}^{pn}$ , but  $p_{i,j}$  is defined on  $\mathbb{R}^p$ .

In some applications we can compute the Gaussian transform  $\langle p_{i,j} \rangle_\lambda$  explicitly. For example, in the distance geometry problem (2.1), the function  $p_{i,j} : \mathbb{R}^p \mapsto \mathbb{R}$  is defined by

$$p_{i,j}(x) = \left( \|x\|^2 - \delta_{i,j}^2 \right)^2. \quad (3.5)$$

This function is decomposable. Moreover, the Gaussian transform is explicitly given by

$$\langle p_{i,j} \rangle_\lambda(x) = p_{i,j}(x) + [3 + (p-1)]\lambda^2 \|x\|^2 + \frac{1}{4}p(p+2)\lambda^4 - p\delta_{i,j}^2\lambda^2. \quad (3.6)$$

For a derivation of this result, see Moré and Wu [35].

As we have noted in Section 2, most of the potentials used in molecular optimization problems are sums of functions of the form  $f(x) = h(\|x\|)$  for some mapping  $h : \mathbb{R} \mapsto \mathbb{R}$ . The following result of Moré and Wu [34] shows that the Gaussian transform of  $f$  can be expressed in terms of one-dimensional integrals.

**Theorem 3.4** *If  $f : \mathbb{R}^3 \mapsto \mathbb{R}$  is of the form  $f(x) = h(\|x\|)$  for some function  $h : \mathbb{R} \mapsto \mathbb{R}$ , then*

$$\langle f \rangle_\lambda(x) = \frac{1}{\lambda\sqrt{\pi}r} \int_0^\infty sh(s) \left[ \exp\left(-\frac{(r-s)^2}{\lambda^2}\right) - \exp\left(-\frac{(r+s)^2}{\lambda^2}\right) \right] ds,$$

where  $r = \|x\|$ . If the mapping  $h$  is an even function, then

$$\langle f \rangle_\lambda(x) = \frac{1}{\lambda\sqrt{\pi}r} \int_{-\infty}^{+\infty} sh(s) \exp\left(-\frac{(r-s)^2}{\lambda^2}\right) ds.$$

Theorem 3.4 can be used to compute the Gaussian transform for molecular optimization problems. As an example, note that the potential function (2.7) used for the confined ionic system has  $h(r) = r^{-1}$ , and thus the Gaussian transform is

$$\begin{aligned} \langle h \rangle_\lambda(x) &= \frac{1}{\lambda\sqrt{\pi}r} \int_0^\infty \left[ \exp\left(-\frac{(r-s)^2}{\lambda^2}\right) - \exp\left(-\frac{(r+s)^2}{\lambda^2}\right) \right] ds \\ &= \frac{1}{\sqrt{\pi}r} \left[ \int_{-\infty}^{r/\lambda} \exp(-t^2) dt - \int_{r/\lambda}^{+\infty} \exp(-t^2) dt \right] \\ &= \frac{1}{\sqrt{\pi}r} \int_{-r/\lambda}^{+r/\lambda} \exp(-t^2) dt = \frac{2}{\sqrt{\pi}r} \text{erf}(r/\lambda), \end{aligned}$$

where  $\text{erf}$  is the standard error function.

Theorem 3.4 reduces the computation of the Gaussian transform to the computation of one-dimensional integrals. We can compute these integrals with standard techniques for numerical integration (for example, an adaptive quadrature), but these techniques usually require a large number of function evaluations. An alternative approach is to use a Gauss-Hermite approximation, as suggested by Moré and Wu [34]. The Gauss-Hermite approximation to the one-dimensional transform

$$\langle f \rangle_\lambda(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} f(x + \lambda s) \exp(-s^2) ds$$

is obtained by using Gaussian quadratures on the integral. Hence,

$$\langle f \rangle_{\lambda,q}(x) = \frac{1}{\sqrt{\pi}} \sum_{i=1}^q w_i f(x + \lambda s_i) \quad (3.7)$$

is the Gauss-Hermite transform, where  $w_i$  and  $s_i$  are, respectively, the standard weights and nodes for Gaussian quadratures. The weights and nodes can be found in the tables of Stroud and Secrest [50] or computed with the `gauss` subroutine in `ORTHOPOL` [11] if the function  $h$  is even. For general  $h$  we should use a Gaussian quadrature for the semi-infinite interval  $[0, +\infty)$ , as advocated by Steen, Byrne, and Gelbard [49].

The Gauss-Hermite transform (3.7) can be viewed as a discrete transform, with carefully chosen sample points. The computation of the Gauss-Hermite transform requires  $q$  function evaluations, but even for modest values of  $q$  we obtain a good approximation to the Gaussian transform.

**Theorem 3.5** *Let  $\langle f \rangle_{\lambda,q}$  be the transformation of  $f : \mathbb{R} \mapsto \mathbb{R}$  defined in (3.7). If  $f^{(l)}$  is piecewise continuous on  $\mathbb{R}$  for some  $l \leq 2q$ , there is a constant  $\mu_l$ , independent of  $f$ , such that*

$$|\langle f \rangle_\lambda(x) - \langle f \rangle_{\lambda,q}(x)| \leq \mu_l \lambda^l \sigma(x),$$

where

$$\sigma(x) = \left\{ \int_{-\infty}^{+\infty} \exp(-s^2) \left| f^{(l)}(x + \lambda s) \right|^2 ds \right\}^{1/2}.$$

Theorem 3.5 is due to Moré and Wu [34]. This result shows that (3.7) is a good approximation to  $\langle f \rangle_\lambda$  provided  $\lambda < 1$ , but that the accuracy is likely to deteriorate if  $\lambda > 1$ . This is not a serious difficulty because for large  $\lambda$  we use (3.7) only to guide an algorithm to a global minimizer, but for small  $\lambda$  we work with the original function  $f$ .

## 4 Computational Experiments

Given the Gaussian transform  $\langle f \rangle_\lambda$ , we can use a continuation algorithm to trace a minimizer of  $\langle f \rangle_\lambda$ . In this section we provide an overview of computational experiments carried out

with a simple continuation algorithm that uses a sequence of continuation parameters

$$\lambda_0 > \lambda_1 > \cdots > \lambda_p = 0.$$

An optimization algorithm is used to determine a minimizer  $x_{k+1}$  of  $\langle f \rangle_{\lambda_k}$ . For  $\lambda_0$  we can use any starting point, but for  $\lambda_k$  with  $k > 0$ , it is reasonable to use  $x_k$  as the starting point. Algorithm **gmin** provides an outline of our continuation algorithm:

**Algorithm gmin**

Choose a random vector  $x_0 \in \mathbb{R}^{m \times 3}$ .

**for**  $k = 0, 1, \dots, p$

Determine  $x_{k+1} = \text{locmin}(\langle f \rangle_{\lambda_k}, x_k)$ .

**end do**

The vector  $x_{p+1}$  is a candidate for the global minimizer. In most cases we use **gmin** with a set of randomly generated starting points with  $p > 0$ . Setting  $p = 0$  in **gmin** reduces to the use of **locmin** on the original function  $f$  from a random starting point. A standard multistart method is obtained if **gmin** is used from a set of randomly generated starting points with  $p = 0$ .

Algorithm **gmin** depends on the optimization procedure **locmin**( $\cdot, \cdot$ ) and on the choice of the continuation parameters  $\lambda_k$ . We will discuss the choice of optimization procedure later; in our computational experiments the continuation parameters are determined by setting

$$\lambda_k = \left(1 - \frac{k}{p}\right) \lambda_0.$$

More sophisticated choices that make use of the behavior of  $\langle f \rangle_{\lambda}$  along the path are clearly possible.

The molecular optimization problems that we are considering can be modeled in terms of the potential function

$$f(x) = \sum_{i,j \in \mathcal{S}} p_{i,j}(x_i - x_j), \quad (4.1)$$

where  $p_{i,j} : \mathbb{R}^3 \mapsto \mathbb{R}$  is the pairwise potential. Algorithm **gmin** can be used to determine the global minimizer of  $f$  once we determine the Gaussian transform  $\langle f \rangle_{\lambda}$ . We have already noted that Theorem 3.3 implies that

$$\langle f \rangle_{\lambda}(x) = \sum_{i,j \in \mathcal{S}} \langle p_{i,j} \rangle_{\sqrt{2}\lambda}(x_i - x_j). \quad (4.2)$$

Hence, we need to determine the Gaussian transform of  $p_{i,j}$ .

For the distance geometry problem (2.1), the pairwise potential  $p_{i,j}$  is defined by (3.5). This potential is decomposable, and the Gaussian transform of  $p_{i,j}$  is given by (3.6). Hence,

(4.2) shows that the Gaussian transform for the distance geometry problem (2.1) is

$$\langle f \rangle_\lambda(x) = \sum_{(i,j) \in \mathcal{S}} \left[ (\|x_i - x_j\|^2 - \delta_{i,j}^2)^2 + 10\lambda^2 \|x_i - x_j\|^2 \right] + \gamma, \quad (4.3)$$

where

$$\gamma = \sum_{(i,j) \in \mathcal{S}} \left( 15\lambda^4 - 6\delta_{i,j}^2 \lambda^2 \right).$$

For the distance geometry problem (2.2), the pairwise potential is

$$p_{i,j}(x) = \min^2 \left\{ \frac{\|x\|^2 - l_{i,j}^2}{l_{i,j}^2}, 0 \right\} + \max^2 \left\{ \frac{\|x\|^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\}.$$

In this case, the potential  $p_{i,j}$  is not decomposable, but  $p_{i,j}(x) = h_{i,j}(\|x\|)$ , where

$$h_{i,j}(r) = \min^2 \left\{ \frac{r^2 - l_{i,j}^2}{l_{i,j}^2}, 0 \right\} + \max^2 \left\{ \frac{r^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\}. \quad (4.4)$$

Since  $h_{i,j}$  is an even function, Theorem 3.4 shows that

$$\begin{aligned} \langle p_{i,j} \rangle_\lambda(x) &= \frac{1}{\lambda \sqrt{\pi} r} \int_{-\infty}^{+\infty} s h_{i,j}(s) \exp \left( -\frac{(r-s)^2}{\lambda^2} \right) ds \\ &= \frac{1}{\sqrt{\pi} r} \int_{-\infty}^{+\infty} (r + \lambda s) h_{i,j}(r + \lambda s) \exp(-s^2) ds, \end{aligned}$$

where  $r = \|x\|$ , and thus (4.2) yields that the Gaussian transform for the distance geometry problem (2.2) is

$$\langle f \rangle_\lambda(x) = \sum_{i,j \in \mathcal{S}} \frac{1}{\sqrt{\pi} r_{i,j}} \int_{-\infty}^{+\infty} (r_{i,j} + \sqrt{2}\lambda s) h_{i,j}(r_{i,j} + \sqrt{2}\lambda s) \exp(-s^2) ds,$$

where  $r_{i,j} = \|x_i - x_j\|$ . In our computational experiments we use the Gauss-Hermite approximation

$$\langle f \rangle_{\lambda,q}(x) = \sum_{i,j \in \mathcal{S}} \frac{1}{\sqrt{\pi} r_{i,j}} \sum_{k=1}^q w_k (r_{i,j} + \sqrt{2}\lambda s_k) h_{i,j}(r_{i,j} + \sqrt{2}\lambda s_k), \quad (4.5)$$

where  $w_k$  and  $s_k$  are the weights and nodes for the Gaussian quadrature, respectively

The functions defined by (4.3) and (4.5) are partially separable because  $\langle f \rangle_\lambda$  and  $\langle f \rangle_{\lambda,q}$  are the sum of  $|\mathcal{S}|$  functions that depend on six variables. We note that the number of flops required to compute the function, gradient, or Hessian matrix of  $\langle f \rangle_\lambda$  and  $\langle f \rangle_{\lambda,q}$  is of order  $|\mathcal{S}|$  because the function and derivatives of each element function can be evaluated with a constant number of flops. An important difference between (4.3) and (4.5), with  $h_{i,j}$  defined by (4.4), is that  $\langle f \rangle_\lambda$  is infinitely differentiable for any  $\lambda \geq 0$ , while  $\langle f \rangle_{\lambda,q}$  is

only continuously differentiable with a piecewise continuous Hessian matrix  $\nabla^2 \langle f \rangle_{\lambda, q}$ . This difference affects the choice of the procedure **locmin**.

In discussing the choice of **locmin**, we assume that we are dealing with distance geometry problems, or more generally, with a problem where  $|\mathcal{S}|$  is bounded by a small multiple of  $n$ . In this case  $\langle f \rangle_\lambda$  and  $\langle f \rangle_{\lambda, q}$  have sparse Hessian matrices, so it makes sense to take advantage of this structure. Problems where  $|\mathcal{S}|$  is essentially  $n^2$  (for example, cluster problems) are discussed in the next section.

A Newton method that takes into account the sparsity of the problem is probably the best choice for **locmin** if the function is defined by (4.3) because for these problems the cost of function, gradient, and Hessian matrix evaluation is of order  $n$ , and the cost per iteration is also of order  $n$ . Moré and Wu [35] used a trust region version because these algorithms are able to escape regions of negative curvature that are present in these problems.

The choice of procedure **locmin** has to be done with some care for the function defined by (4.5) because  $\langle f \rangle_{\lambda, q}$  is not twice continuously differentiable. The Hessian matrix is discontinuous at points where the argument of  $h_{i,j}$  coincides with either  $l_{i,j}$  or  $u_{i,j}$ . We cannot expect to avoid these discontinuities, in particular, if  $l_{i,j}$  or  $u_{i,j}$  are close. Moré and Wu [34] used the variable-metric limited-memory code **vmlm** in **MINPACK-2**, which is an implementation of the Liu and Nocedal [33] algorithm.

The formulation of the distance geometry problem (2.2) in terms of  $p_{i,j}(x) = h_{i,j}(\|x\|)$  where  $h_{i,j}$  is defined by (4.4) is typical, but other formulations have been used. Crippen and Havel [9] used

$$h_{i,j}(r) = \min^2 \left\{ \frac{r^2 - l_{i,j}^2}{r^2}, 0 \right\} + \max^2 \left\{ \frac{r^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\},$$

because they felt that this formulation lead to a problem with fewer minimizers, but Havel [20] advocates the use of

$$h_{i,j}(r) = \min^2 \left\{ \frac{r^2 - l_{i,j}^2}{r^2 + l_{i,j}^2}, 0 \right\} + \max^2 \left\{ \frac{r^2 - u_{i,j}^2}{u_{i,j}^2}, 0 \right\}$$

because this formulation avoids the barrier created at  $r = 0$ . In both formulations  $p_{i,j}$  has a discontinuous second derivative. If we use

$$h_{i,j}(r) = \begin{cases} \left( \frac{l_{i,j}^2 - r^2}{l_{i,j}^2} \right)^3 & \text{if } r < l_{i,j}, \\ 0 & \text{if } r \in [l_{i,j}, u_{i,j}], \\ \left( \frac{r^2 - u_{i,j}^2}{u_{i,j}^2} \right)^3 & \text{if } r > u_{i,j}, \end{cases} \quad (4.6)$$

then  $p_{i,j}$  is twice continuously differentiable. Moreover, if  $f$  is defined by (4.1), then  $f(x) \geq 0$  with  $f(x) = 0$  if and only if  $x$  solves the distance geometry problem (2.2). From an



optimization point of view, formulation (4.6) is preferable because it allows the use of a Newton method in `locmin`.

The computational experiments performed by Moré and Wu [35, 34] on various distance geometry problems show that algorithm `gmin` is able to find global minimizers reliably and efficiently. An interesting aspect of these results is that algorithm `gmin` with  $p > 0$  requires less than twice the effort (measured in terms of function and gradient evaluations) than  $p = 0$ . At first sight this is surprising because `gmin` with  $p > 0$  requires the solution of  $p$  minimization problems. However, for reasonable choices of  $\lambda_0$ , finding a minimizer of  $\langle f \rangle_\lambda$  with  $\lambda = \lambda_0$  is found quickly because  $\langle f \rangle_\lambda$  is a smooth, well-behaved function. A minimizer of  $\langle f \rangle_\lambda$  with  $\lambda = \lambda_k$  is also found quickly because  $x_k$  is a good starting point. On the other hand, `gmin` with  $p = 0$  must find a local minimizer of  $f$ , which is not necessarily smooth or well-behaved, from a starting point that is not guaranteed to be near a minimizer. We expect that future work will improve the continuation procedure and further reduce the cost of the continuation procedure.

## 5 Future Directions

Global smoothing and continuation have proved to be effective tools for the solution of molecular optimization problems with a moderate number of atoms, but improvements in these techniques will be needed to address problems with a large number of atoms. In this section we outline possible extensions to the work that we have presented.

### 5.1 Continuation Algorithms

Algorithm `gmin` is a relatively simple algorithm for tracing a curve  $x(\lambda)$ , where  $x(\lambda)$  is a minimizer of  $\langle f \rangle_\lambda$ . For problems with a large number of atoms we need to improve `gmin` by computing  $x(\lambda)$  more efficiently. If we define function  $h : \mathbb{R}^n \times \mathbb{R} \mapsto \mathbb{R}$  by

$$h(x, \lambda) = \langle f \rangle_\lambda(x),$$

and differentiate twice with respect to the variable  $x$ , we obtain

$$\partial_{xx}h[x(\lambda), \lambda]x'(\lambda) + \partial_{\lambda x}h[x(\lambda), \lambda] = 0.$$

This differential equation, together with the initial value  $x(0) = x_0$ , defines a trajectory  $x(\lambda)$  under suitable nondegeneracy assumptions. We can use continuation algorithms (see, for example, Allgower and Georg [2]), but these algorithms are designed to trace stationary points of  $h$ , that is, solutions to

$$\partial_x h[x(\lambda), \lambda] = 0.$$

Our situation is somewhat different because we need to trace minimizers of  $h$ . However, in general it is not possible to define a continuous trajectory of minimizers, and thus we

must be prepared to jump curves. For additional information on issues related to tracing minimizers, see Gudat, Guerra Vazquez, and Jongen [17].

## 5.2 Smoothing

The Gaussian transform is isotropic because if we view the function  $f$  in a different coordinate system via the function  $h : \mathbb{R}^n \mapsto \mathbb{R}$  defined by

$$h(x) = f(P^T x),$$

then  $\langle h \rangle_\lambda(Px) = \langle f \rangle_\lambda(x)$  for any orthogonal matrix  $P \in \mathbb{R}^{n \times n}$ . If we wish to emphasize some directions, then Wu [51] suggested the use of the *anisotropic* Gaussian transform defined by

$$\langle f \rangle_\Lambda(x) = \frac{1}{\pi^{n/2} |\det \Lambda|} \int_{\mathbb{R}^n} f(y) \exp\left(-\|\Lambda^{-1}(y - x)\|^2\right) dy$$

for any nonsingular matrix  $\Lambda$ . For this transformation  $\langle h \rangle_\Lambda(Px) = \langle f \rangle_\Lambda(x)$  if  $P\Lambda$  is orthogonal, so that the scaling in  $\Lambda$  controls the smoothing. Wu [51] showed that if  $\Lambda$  is a diagonal matrix, then this transformation can be used for decomposable functions and potential functions in molecular optimization problems.

The Gaussian transform can be extended to a general density function  $\rho : \mathbb{R}^n \mapsto \mathbb{R}$  by defining the *generalized* transform by

$$\langle\langle f \rangle\rangle_\lambda(x) = \frac{1}{\lambda^n} \int_{\mathbb{R}^n} f(y) \rho\left(\frac{x - y}{\lambda}\right) dy.$$

The analysis of Moré and Wu [36] shows that the smoothing properties of the Gaussian transform can be extended to this class of transformation. By admitting a larger class of transformations, we should be able to extend the range of functions that can be transformed.

## 5.3 Optimization Algorithms

Newton methods are appropriate for distance geometry problems where  $|\mathcal{S}|$  is of order  $n$  because for these problems the cost of function, gradient, and Hessian matrix evaluation is of order  $n$ , and the cost per iteration is also of order  $n$ . A standard Newton method is not appropriate for large cluster problems where  $|\mathcal{S}|$  is of order  $n^2$  because the storage is of order  $n^2$  and the cost per iteration is of order  $n^3$ .

The limited-memory variable-metric method [33] is suitable for systems with a large number of atoms because the memory requirements and cost per iteration is of order  $n$ . Unfortunately, the number of iterations required for convergence on these problems increases rapidly with the number of atoms. Preliminary experiments with a standard truncated Newton method showed that this method required a large number of inner conjugate gradient iterations. Since each conjugate gradient iteration requires order  $n$  operations, it is

not surprising that the standard truncated Newton method required more computing time than the limited-memory variable-metric method.

We expect that a truncated Newton method with a suitable preconditioner will reduce the computing time required to solve cluster problems. Schlick and Fogelson [45, 46] developed such an algorithm for molecular dynamics simulation and structure refinement, with a preconditioner constructed from an approximate Hessian matrix. Similar ideas should apply to cluster problems.

#### 5.4 Function Evaluations

We have already noted that the number of flops required to compute the function and derivatives in a distance geometry problem is of order  $|\mathcal{S}|$ . In distance geometry problems  $|\mathcal{S}|$  is of order  $n$ , and thus we can evaluate these functions in order  $n$ . In cluster problems, however, all pairwise potentials are included, and then the cost of evaluating the function, gradient, and Hessian matrix is of order  $n^2$ . This represents a major hurdle to the solution of large cluster problems, since in a typical problem we need multiple runs and hundreds of function evaluations per run to determine the global minimizer.

We can reduce the cost of the function evaluation by computing an approximation to the function. The fast multipole method (Greengard and Rokhlin [15], and Greengard [14]), in particular, has attracted considerable attention because the running time is proportional to  $n$ . However, implementation of the fast multipole method requires considerable care and analysis, so only sophisticated implementations are able to achieve the order  $n$  running time.

Board and coworkers [32, 31, 40] have developed several sequential and parallel packages for computing electrostatic force fields and potentials using fast multipole algorithms. These implementations have been done with considerable care but are geared to molecular dynamics simulations where it is reasonable to assume a uniform distribution of atoms. The performance of these algorithms degrades considerably in an optimization setting because the distribution of the atoms is not uniform, unless we are in the final stages of convergence. In our opinion the only currently effective method for reducing the computing time of the function evaluation in large cluster problems is to evaluate the function in parallel.

#### 5.5 High-Performance Architectures

Macromolecular systems usually have 1,000 to 10,000 atoms, and the ionic systems of physical interest contain up to 200,000 ions. Determining the global solutions for these problems will not be feasible without the use of parallel high-performance architectures, even with the most efficient optimization algorithm.

The global continuation algorithm can be parallelized easily at a coarse level, with each processor assigned the computation of a solution trajectory. This strategy requires little

communication among processors and is suitable for massively parallel architectures, such as the IBM SP. Indeed, we have implemented, for example, the continuation algorithms for distance geometry problems on the IBM SP at Argonne. Although we have not yet tested the algorithms with large problems, the results on medium-sized problems (with 500 atoms) show that the algorithms have satisfactory performance on as many as 64 processors.

Load balancing and synchronization between the processors are two of the problems that must be addressed for systems with a large number of atoms. Load balancing can be a problem because trajectories may require different amount of computing time. Another problem is that processors may end up tracing the same solution trajectory, even if they are given different starting points. Synchronization between the processors will be required to make sure that different trajectories are traced.

## Acknowledgments

Our research on macromolecular global optimization problems has been influenced by John Schiffer's work on ionic systems and Julius Jellinek's work on clusters. Steve Pieper deserves special mention for bringing the work on ionic systems to our attention and for sharing his insights on this problem. We are also grateful to Ian Coope for pointing out that the distance geometry problem (2.1) can be solved as a linear algebra problem if all the pairwise distances are available. Gail Pieper's incisive comments on the manuscript are also gratefully acknowledged.

## References

- [1] S. AL-HOMIDAN AND R. FLETCHER, *Hybrid methods for finding the nearest Euclidean distance matrix*, technical report, The University of Dundee, Dundee, Scotland, 1995.
- [2] E. L. ALLGOWER AND K. GEORG, *Numerical Continuation: An Introduction*, Springer-Verlag, 1990.
- [3] L. M. BLUMENTHAL, *Theory and Applications of Distance Geometry*, Oxford University Press, 1953.
- [4] A. T. BRÜNGER AND M. NILGES, *Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy*, Q. Rev. Biophys., 26 (1993), pp. 49–125.
- [5] R. H. BYRD, E. ESKOW, AND R. B. SCHNABEL, *A new large-scale global optimization method and its application to Lennard-Jones problems*, Technical report CU-CS-630-92, Department of Computer Science, University of Colorado, Boulder, Colorado, revised, 1995.

- [6] R. H. BYRD, E. ESKOW, R. B. SCHNABEL, AND S. L. SMITH, *Parallel global optimization: Numerical methods, dynamic scheduling methods, and application to molecular configuration*, in *Parallel Computation*, B. Ford and A. Fincham, eds., Oxford University Press, 1993, pp. 187–207.
- [7] T. F. COLEMAN, D. SHALLOWAY, AND Z. WU, *Isotropic effective energy simulated annealing searches for low energy molecular cluster states*, *Comp. Optim. Applications*, 2 (1993), pp. 145–170.
- [8] ———, *A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing*, *J. Global Optim.*, 4 (1994), pp. 171–185.
- [9] G. M. CRIPPEN AND T. F. HAVEL, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, 1988.
- [10] M. R. GAREY AND D. S. JOHNSON, *Computers and Intractability*, W. H. Freeman, 1979.
- [11] W. GAUTSCHI, *Algorithm 726: ORTHOPOL – A package of routines for generating orthogonal polynomials and Gauss-type quadrature rules*, *ACM Trans. Math. Software*, 20 (1994), pp. 21–62.
- [12] W. GLUNT, T. L. HAYDEN, AND M. RAYDAN, *Molecular conformation from distance matrices*, *J. Comp. Chem.*, 14 (1993), pp. 114–120.
- [13] ———, *Preconditioners for distance matrix algorithms*, *J. Comp. Chem.*, 15 (1994), pp. 227–232.
- [14] L. GREENGARD, *The Rapid Evaluation of Potential Fields in Particle Systems*, MIT Press, 1988.
- [15] L. GREENGARD AND V. ROKHLIN, *A fast algorithm for particle simulation*, *J. Comput. Phys.*, 73 (1987), pp. 325–348.
- [16] A. GRIEWANK, *Generalized descent for global optimization*, *J. Optim. Theory Appl.*, 34 (1981), pp. 11–39.
- [17] J. GUDDAT, F. G. VAZQUEZ, AND H. T. JONGEN, *Parametric Optimization: Singularities, Pathfollowing and Jumps*, John Wiley & Sons, 1990.
- [18] H. HABERLAND, ed., *Clusters of Atoms and Molecules*, *Springer Series in Chemical Physics*, vol. 52, Springer-Verlag, 1994.
- [19] R. W. HASSE AND J. P. SCHIFFER, *The structure of the cylindrically confined coulomb lattice*, *Ann. Physics*, 203 (1990), pp. 419–448.

- [20] T. F. HAVEL, *An evaluation of computational strategies for use in the determination of protein structure from distance geometry constraints obtained by nuclear magnetic resonance*, Prog. Biophys. Mol. Biol., 56 (1991), pp. 43–78.
- [21] B. A. HENDRICKSON, *The molecule problem: Determining conformation from pairwise distances*, PhD thesis, Cornell University, Ithaca, New York, 1991.
- [22] ———, *The molecule problem: Exploiting structure in global optimization*, SIAM J. Optimization, 5 (1995), pp. 835–857.
- [23] M. R. HOARE, *Structure and dynamics of simple microclusters*, Advances in Chemical Physics, 40 (1979), pp. 49–135.
- [24] M. R. HOARE AND J. MCINNES, *Statistical mechanics and morphology of very small atomic clusters*, Faraday Discuss. Chem. Soc., 61 (1976), pp. 12–24.
- [25] M. R. HOARE AND P. PAL, *Statistics and stability of small assemblies of atoms*, J. Cryst. Growth, 17 (1972), pp. 77–96.
- [26] J. JELLINEK, *Theoretical dynamical studies of metal clusters and cluster-ligand systems*, in Metal-Ligand Interactions: Structure and Reactivity, N. Russo, ed., Kluwer Academic Publishers, 1995 (in press).
- [27] J. KOSTROWICKI AND L. PIELA, *Diffusion equation method of global minimization: Performance for standard functions*, J. Optim. Theory Appl., 69 (1991), pp. 269–284.
- [28] J. KOSTROWICKI, L. PIELA, B. J. CHERAYIL, AND H. A. SCHERAGA, *Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms*, J. Phys. Chem., 95 (1991), pp. 4113–4119.
- [29] J. KOSTROWICKI AND H. A. SCHERAGA, *Application of the diffusion equation method for global optimization to oligopeptides*, J. Phys. Chem., 96 (1992), pp. 7442–7449.
- [30] I. D. KUNTZ, J. F. THOMASON, AND C. M. OSHIRO, *Distance geometry*, in Methods in Enzymology, N. J. Oppenheimer and T. L. James, eds., vol. 177, Academic Press, 1993, pp. 159–204.
- [31] C. G. LAMBERT AND J. A. BOARD, *A multipole-based algorithm for efficient calculation of forces and potentials in macroscopic periodic assemblies of particles*, Technical report 95-001, Department of Electrical Engineering, Duke University, Durham, North Carolina, 1995.
- [32] J. F. LEATHRUM AND J. A. BOARD, *The parallel fast multipole algorithm in three dimensions*, Technical report TR92-001, Department of Electrical Engineering, Duke University, Durham, North Carolina, 1992.

- [33] D. C. LIU AND J. NOCEDAL, *On the limited memory BFGS method for large scale optimization*, Math. Programming, 45 (1989), pp. 503–528.
- [34] J. J. MORÉ AND Z. WU,  *$\varepsilon$ -optimal solutions to distance geometry problems via global continuation*, in Global Minimization of Nonconvex Energy Functions: Molecular Conformation and Protein Folding, P. M. Pardalos, D. Shalloway, and G. Xue, eds., American Mathematical Society, 1995, pp. 151–168.
- [35] ———, *Global continuation for distance geometry problems*, Preprint MCS-P505-0395, Argonne National Laboratory, Argonne, Illinois, 1995.
- [36] ———, *Smoothing techniques for macromolecular global optimization*, Preprint MCS-P542-0995, Argonne National Laboratory, Argonne, Illinois, 1995.
- [37] J. A. NORTHBY, *Structure and binding of Lennard-Jones clusters:  $13 \leq n \leq 147$* , Journal of Chemical Physics, 87 (1987), pp. 6166–6177.
- [38] L. PIELA, J. KOSTROWICKI, AND H. A. SCHERAGA, *The multiple-minima problem in the conformational analysis of molecules: Deformation of the protein energy hypersurface by the diffusion equation method*, J. Phys. Chem., 93 (1989), pp. 3339–3346.
- [39] R. RAFAC, J. P. SCHIFFER, J. S. HANGST, D. H. E. DUBIN, AND D. J. WALES, *Stable configurations of confined cold ionic systems*, Proc. Natl. Acad. Sci. U.S.A., 88 (1991), pp. 483–486.
- [40] W. T. RANKIN AND J. A. BOARD, *A portable distributed implementation of the parallel multipole tree algorithm*, Technical report 95-002, Department of Electrical Engineering, Duke University, Durham, North Carolina, 1995.
- [41] P. J. REYNOLDS, ed., *On Clusters and Clustering*, North-Holland, 1993.
- [42] J. B. SAXE, *Embeddability of weighted graphs in  $k$ -space is strongly NP-hard*, in Proc. 17th Allerton Conference in Communications, Control and Computing, 1979, pp. 480–489.
- [43] H. A. SCHERAGA, *Predicting three-dimensional structures of oligopeptides*, in Reviews in Computational Chemistry, K. B. Lipkowitz and D. B. Boyd, eds., vol. 3, VCH Publishers, 1992, pp. 73–142.
- [44] J. P. SCHIFFER, *Phase transitions in anisotropically confined ionic crystals*, Phys. Rev. Lett., 70 (1993), pp. 818–821.
- [45] T. SCHLICK AND A. FOGELSON, *TNPACK – A truncated Newton minimization package for large-scale problems: I. Algorithms and usage*, ACM Trans. Math. Software, 18 (1992), pp. 46–70.

- [46] ———, *TNPACK – A truncated Newton minimization package for large-scale problems: II. Implementations examples*, ACM Trans. Math. Software, 18 (1992), pp. 71–111.
- [47] D. SHALLOWAY, *Application of the renormalization group to deterministic global minimization of molecular conformation energy functions*, J. Global Optim., 2 (1992), pp. 281–311.
- [48] ———, *Packet annealing: A deterministic method for global minimization, application to molecular conformation*, in Recent Advances in Global Optimization, C. Floudas and P. Pardalos, eds., Princeton University Press, 1992, pp. 433–477.
- [49] N. M. STEEN, G. D. BYRNE, AND E. M. GELBARD, *Gaussian quadratures for the integrals  $\int_0^\infty \exp(-x^2)f(x)dx$  and  $\int_0^b \exp(-x^2)f(x)dx$* , Math. Comp., 23 (1969), pp. 661–674.
- [50] A. H. STROUD AND D. SECREST, *Gaussian Quadrature Formulas*, Prentice-Hall, Inc., 1966.
- [51] Z. WU, *The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation*, Preprint MCS-P442-0694, Argonne National Laboratory, Argonne, Illinois, 1994.
- [52] G. L. XUE, *Improvement on the Northby algorithm for molecular conformation: Better solutions*, J. Global. Optim., 4 (1994), pp. 425–440.